

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 1034.21 Seconds  
(without alignments)  
1371.762 Million cell updates/sec

Title: US-09-786-502A-8  
Perfect score: 30  
Sequence: 1 tcgagatcttgcaggacgataggctgc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.cm.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.roi.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	30	100.0	30	6	BD272285 Fused rec
2	21	70.0	10951	1	AE005756 Caulobact
3	20.4	68.0	668	9	AF344837 Papio cyn
4	20.4	68.0	668	9	AF344842 Cercocebu
5	20.4	68.0	668	9	AF344845 Callithri
6	20.4	68.0	668	9	AF344852 Macaca ne
7	20.4	68.0	668	9	AF344855 Macaca mu
8	20.4	68.0	875	6	AR102241 Sequence
9	20.4	68.0	1514	6	BD267803 Methods a
10	20.4	68.0	1514	6	I07254 Sequence 11
11	20.4	68.0	1514	6	AR380771 Sequence
12	20.4	68.0	1514	6	BD015187 TtiSa cel
13	20.4	68.0	1514	9	J02988 Human T-cel
14	20.4	68.0	2405	6	AR102243 Sequence
15	20.4	68.0	3230	9	HUMCD284
16	20.4	68.0	3585	6	AR102242 Sequence
17	20.4	68.0	3803	6	AX557214 Sequence
18	20.4	68.0	3804	6	AX695882 Sequence
19	20.4	68.0	3806	6	CQ715869 Sequence

C 20	20.4	68.0	3806	6	AX283559 Sequence
C 21	20.4	68.0	51365	6	AX695881 Sequence
C 22	20.4	68.0	106539	9	AF225899 Homo sapi
C 23	20.4	68.0	196622	9	AC125238 Homo sapi
C 24	20.2	67.3	181025	10	AC147243 Mus muscu
C 25	20.2	67.3	206894	2	AC123799 Mus muscu
C 26	20.2	67.3	216521	10	AC110230 Mus muscu
C 27	20	66.7	142422	9	AC113935 Homo sapi
C 28	20	66.7	158498	2	AC098295 Rattus no
C 29	20	66.7	163291	2	AC102390 Mus muscu
C 30	20	66.7	172613	9	AC093433 Homo sapi
C 31	20	66.7	173995	2	AC073027 Homo sapi
C 32	20	66.7	203812	2	AC026874 Homo sapi
C 33	20	66.7	232735	2	AC115240 Rattus no
C 34	20	66.7	237445	2	AC120701 Rattus no
C 35	20	66.7	243439	2	AC129059 Rattus no
C 36	20	66.7	280288	2	AC127373 Mus muscu
C 37	20	66.7	300050	1	AP004596 Oceanobac
C 38	20	66.7	312728	2	BX005159 Homo sapi
C 39	20	66.7	331221	2	AC136270 Rattus no
C 40	19.6	65.3	5359	8	AY057108 Hypocrea
C 41	19.6	65.3	149121	5	BX545918 Zebrafish
C 42	19.6	65.3	194224	5	AL954868 Zebrafish
C 43	19.4	64.7	30	6	AL63752 Sequence 16
C 44	19.4	64.7	64	6	A63755 Sequence 19
C 45	19.4	64.7	852	6	AX652869 Sequence

ALIGNMENTS

RESULT 1  
BD272285  
LOCUS  
DEFINITION Fused receptor specific to prostate-specific membrane antigen and its use.  
ACCESSION BD272285  
VERSION BD272285.1 GI:33082053  
KEYWORDS JP 2002524081-A/8.  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Sadelain,M., Bander,N.H. and Gong,M.  
TITLE Fused receptor specific to prostate-specific membrane antigen and its use  
JOURNAL Patent: JP 2002524081-A 8 06-AUG-2002;  
SLOAN KETTERING INSTITUTE FOR CANCER RESEARCH  
COMMENT OS Synthetic construct  
PN JP 2002524081-A/8  
PD 06-AUG-2002  
PF 03-SEP-1999 JP 2000568998  
PI MICHEL SADELAIN,NEIL H BANDER,MICHAEL GONG  
PC C12N15/09,A61K35/26,A61K39/395,A61K48/00,A61P35/00,  
PC C07K19/00,  
CC C12N5/10//C07K16/44,C12N15/00,C12N5/00  
CC Fused receptor specific to prostate-specific membrane antigen and its use  
FH Key Location/Qualifiers  
FT source 1..30  
FT Location/Qualifiers  
source 1..30  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

ORIGIN

Query Match 100.0%; Score 30; DB 6; Length 30;  
Best Local Similarity 100.0%; Pred. No. 0.0016;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
Db |||||
1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30

RESULT 2
AE005756 1951 bp DNA linear BCT 12-JUN-2002
DEFINITION Caulobacter crescentus CB15 section 82 of 359 of the complete genome.
ACCESSION AE005756
VERSION AE005756.1
KEYWORDS GI:13422042
SOURCE
ORGANISM
Caulobacter crescentus CB15
Caulobacter crescentus CB15
Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
Caulobacteraceae; Caulobacter.
1 (bases 1 to 10951)
Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
Shetty,J., Berry,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
11259647
2 (bases 1 to 10951)
Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
Berry,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
Fraser,C.M.
Direct Submission
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 10951
/organism="Caulobacter crescentus CB15"
/mol type="genomic DNA"
/strain="CB15"
/db xref="taxon:190650"
177..2102
/genes="CC0798"
177..2102
/genes="CC0798"
/notes="similar to GB:U40408 PID:1373055 PID:1628565
PID:1668766; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sialic acid-specific 9-O-acetyltransferase,
putative"
/protein_id="AAK22783.1"
/db xref="GI:13422043"
/transl_table=11
/translation="MSLKALILATMGLSTPALAGPLASVFSHAVLQRDQIRVWGR
AAPNVAVSLSGGVYASADGRWSAVLPARPAGGALTTLTARAGDSTQTVSNLVM
GDVNLCSGYSYRQALNGDGEVYASADGRDRIILQTKISRPFPQSLPAGVYWK
VSTPQAAAFSAACVFMGRRLRTTGVIGLIDATWGGSVIQDWISREGLAALKTYDE
GLSVLDYADPALCVFMGRWSQSLRWADAKI PAAGKWEKPPDDRAWKMPTEGFWEO
AAPDLAPDGTWMLRLTLTKAAQAAGATLALGPVDDIDITFLNGREIGSTQWDTTP
RTYRLAPGALGVSNVLAIRVIMDGGGSPGWCKAALKGLTADGTVPVLTATYRYKIA
APLSELTAPSPALGSGVGLTUNGRIAPLVFVGKFAWYGEANVTEAQARYKIA
PAMIAWRARAFGNGQPFLLTQLAFAFPEVDREPVNSAAALRDVQVAAADPMVGM
SAVDVGSYPDIDHPADKARVQRLQALAYGEAAVAAAPLPSRREGETVWVTL
QPAVWAGNRPICIFELCDAAACFVDATVEGGSVRLATPAGMTATKRVAVADSPIN
LFCANRLPATPFELVIP"
complement (2112..3659)

/genes="CC0799"
complement (2112..3659)
/genes="CC0799"
/notes="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="para-nitrobenzyl esterase"
/protein_id="AAK22784.1"
/db xref="GI:13422044"
/translation="MGFTAEARSPPVATTNGKRGVLDGEVSFKGLRGYADTGAGARR
FWPPVKPEPTEVKDALAYGPASMTQKGEGETLSEDCLEFNVTWTPARSKTGLAD
GAKRPMFYHIGAYNGSGSPWEGTKLARGDVVYVTVNHLNKAFTLYLWMPGSAK
APSDASNGVQLDLVLALQWVRDNIAREGGDDPCVNLFGQGGGAKIATLWMPGSAK
GLFRAATVSGQVTVGGPFNATRAKAFDLKLGVDKLAALRALPAEMLAGLKAYDP
IAGSGVYVGVLDQRLSLRHPFFDPAAPQSLSPMVGNTHDTEKFGIOWDAKAPQ
TWDEIARLPQGAARIDIDPTVFAFYQTYPNYSPADVFFAASAGRSWKAIIQD
EERAKAPAFAYQVNWRSPIOGGIFGAPHTIDIGLVECTLDAGKSIVCTGPDVSAMS
NTMSDAFIAPARTGDPNGALPKWEPYTLPRRQTWVFDIVSRLEDDPRGVEREFNRV
PTQGT"
complement (3818..5542)
/genes="CC0800"
complement (3818..5542)
/genes="CC0800"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22785.1"
/db xref="GI:13422045"
/translation="NAWRPMIQAATWMAATLALGVPVAALGCSPPVQGVYVPSNFELVOI
ADAVVAEAGVNRASDEDFRKKVVFVRAQPKGAGNEIEVSSVUSVKIQSPDPE
ILIEPHEVARGACNRTIVAKGTYVLFLKVEDSYASLGYPFSRINVEDYAGEASPMWL
TIRTLRLQNSQPMQATLERMQAELIAKTPTRGEQAQAMADLHILGAPSPWKPT
AYLLAAYDDVKAQRPYKPRSTYDAERSEAAFTDLIMENAGLSDGKPPPPRRDP
RDLALLEWMDGHEPGAMEPLDALAASDAPGGDLAMAIRFYAKNGRYDAYGLIETRA
APLTLPEGEAWTLIDREAOQGEYDGRPRWRSOPDLAARWPRLAVALSRITQD
RGRDADFVSLMSLGSYRNPBLTMLSRSNDIIIAWAERELVAPANLTASSVGP
DFLKLPMKVIILRWMLGDDKDVAVLTMPFCRGAQRMQFMFDITGLSTGDAHALL
RFAASSVMTDSRQALAPASIAWDKRYANESNMWQTDPAWRKAANGQPIATAKDIMP
LKPVSQAS"
complement (5626..6954)
/genes="CC0801"
complement (5626..6954)
/genes="CC0801"
/notes="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="mannanase, putative"
/protein_id="AAK22786.1"
/db xref="GI:13422046"
/translation="MTLSRRHLIAAAALPAIGAAKAPTPPKGVTVTKDGRSLDGKPK
YRFAGTNVYAAWLGAPAGYGLRLRLRELDRLKAMGVNLRILGAGEQSPAKVAMD
TRFGPGEYVADLLKGLDVTLAEMQRMKAVIYVNNFWDGSGMPAYLWNVGDGPF
QOQDPAYPWPQYADYSARFYANKALFRHYVTSLSVRSVTYTKGPRDDPTIMSWQ
LANFRPGSDAFQGSNMPAYAWIRDTAGLIKRLDNHLSVTGSGTGWCGLESQV
IDAHPVVIDYMTLHIWPNWGWISMTDQPSYEAGEQKCRDYVAQHIALAKRLKPL
TIEBFGLIRDGQFTGPTTDRFYKTMLDLADLMKAGGPTAGVNFWMNGEGR
QOQDAWFKKDKSYVGDPPQEEQGLGVFDADASTLAVIKAHAAAVKALG"
complement (7030..7311)
/genes="CC0802"
complement (7030..7311)
/genes="CC0802"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK22787.1"
/db xref="GI:13422047"
/translation="MTFTLVSVRAKRDNFRLIIVMLVERDPRAAARLGLPILRALDSL
TEAPRGRSGVGTTRTIEISIPFGQSAYVIRYLLGSSVHVTRIWHGLEQR"
complement (7308..7571)
/genes="CC0803"
complement (7308..7571)

```

```

/gene="CC0803"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22788.1"
/db_xref="GI:13422048"
/transl_table="MADGFDIHDQEQARLKVAVDRLGMSVSEYAVALIDAGLTGAA"
PKAIDPPAIDEAIADEAIEDGDEFAISRDEFAHRRVVTAGL"
complement(7610..10699)
/gene="CC0804"
complement(7610..10699)
/gene="CC0804"
/note="identified by match to PFAM protein family HMM
PF01619"
/codon_start=1
/transl_table=11
/product="proline
dehydrogenase/delta-1-pyrroline-5-carboxylate
dehydrogenase"
/protein_id="AAK22789.1"
/db_xref="GI:13422049"
/transl_table="MTDMSLDAGKYEDEAAVIADLLAAKPLSEDEAAVRAEAALV
RGARSVRKGVSEFLQESLGTREGALMCLAEALLRTPDDTRDKLIAEKIGSAD
WASHLGGSDLSLVNASTGWLMLTKIVPEDETARNDMPGFIKKLAGRLGPEPVARAAV
QAIRIMGEQVGLRTIEAAILKRAAEGDMGCFDMLGEGARTAAADARYEKAYADALET
VGLSINGAGPEAGHVSYSALSCPRYEATHEDRVWELYPRTLRLAKIAARNLNFT
IDAEADRLALSLLKLDKCREPELGDWTGLGLAVOAYOKRCGCEVARTARLKAASEETGR
RLMVLVKGAYWSEIKRAQVAGRPDPVPTTKPATDLSLVNAKALIEAPHLYAQP
ATNNAVLAAVMAKNTGVKIEHQRLHGMGEALYKAADDLYDGTITURATAPVGGHED
LLPLVRRLLLENGANTS FVHALLDERVPVEKVTDPIDTVEAHPDRHAKIPTPINVQ
LERNVAGSLDLSVKARERISAAVAQDGTLSAGPLVGGKVAGGAPLPIIAPANDQ
KTVGVSEAGSQAIDEAFKLARAQAQPAWDRAQGVARAQVLRMGDALENIETLIALL
SRAGKTLDSGIEAEVRAVDFCRYAMLAEDQGEAEILKAGPVGNTSLRLAGRGVVF
CISPPNPPLAIFTQIQAALAAAGNVLAKPAEQTPLIAFAVVKLYHNAGLDLPRIALL
PGRTEGAAATSHEDDVGAFCTGTATWINTQTLAARQGPVFPVFAETGGNGMFPV
DTTAQREQVDDVVISAFSGAGQCSALRLLFLPHPTADHIIIEGLGAMDALVIGDPA
LVFDGVPVLDIAKALDHLVRLKSDAKVLHAAAPAGCTFPFAPVLAIEPTADFL
REVFGPVHVRYKPNLEKVGALAAARRYGLTGLIHRSIESFAADVQRLVAPGNAVYV
NRSMTGAVVGVQPPGGBLSGTGPKAGGPHALLRFAVERALS VYNITAQGGDPLLNL"

ORIGIN
Query Match      70.0%; Score 21; DB 1; Length 10951;
Best Local Similarity 82.8%; Pred. No. 75;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TCGAGGATCTGTCAGGAGCGATAGGCTG 29
Db 5229 TCAGGATCGTCAGGATCGCTGGCTG 5257

RESULT 3
AF344837/c
LOCUS
DEFINITION
Papio cynocephalus anubis CD28 protein precursor, mRNA, complete
cde.
AF344837
VERSION AF344837.1 GI:13649990
KEYWORDS
SOURCE
ORGANISM
Papio anubis (olive baboon)
Papio anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopithecinae; Papio.
1 (bases 1 to 668)
Vallinger,F., Bostik,P., Mayne,A., King,C.L., Genain,C.P.,
Weiss,W.R. and Ansari,A.A.
Cloning, sequencing, and homology analysis of nonhuman primate
Fas/Fas-ligand and co-stimulatory molecules
Immunogenetics 53 (4), 315-328 (2001)
21383618
PUBMED 11491535
REFERENCE 2 (bases 1 to 668)
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

```

AUTHORS Villinger, F.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-FEB-2001) Pathology and Laboratory Medicine, Emory University School of Medicine, Winship Cancer Institute, 1365B Clifton Rd, Atlanta, GA 30322, USA  
 FEATURES Location/Qualifiers  
 source 1. .668  
 /organism="Papio anubis"  
 /mol\_type="mRNA"  
 /sub\_species="anubis"  
 /db\_xref="taxon:9555"  
 CDS 2. .664  
 /note="co-stimulatory molecule"  
 /codon\_start=1  
 /product="CD28 protein precursor"  
 /protein\_id="AAK37533.1"  
 /db\_xref="GI:13649991"  
 /translation="MLRLLALNLLPSIQVTGNKILVKQSPMLVAYDNAVNLCKYSY  
 NLFPSRFRASLHKGLDSAVEVCVYGNYSQQLQYISKTGFCGCKLGDSVTFYLQNM  
 YNQTDIYCKIEVMYPPPLYDKNGKNTIIHVKGKHLCPSPFPGSGKPFWALVVVGN  
 GLVACYSLLVTTVAFSIFCMRSKRSLHSDYNNMTPRPGTRKHQPYAPPRDFAAY  
 RS"  
 sig\_peptide 2. .55  
 ORIGIN  
 Query Match 68.0%; Score 20.4; DB 9; Length 668;  
 Best Local Similarity 95.5%; Pred. No. 1.1e+02;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 9 CTTGTCAGGAGCGATAGGCTGC 30  
 | ||||| ||||| ||||| |||||  
 Db 668 CGTGTCCAGGAGCGATAGGCTGC 647  
 RESULT 4  
 AF344842/c  
 LOCUS Cercocebus torquatus atys (sooty mangabey)  
 DEFINITION Cercocebus torquatus atys  
 ACCESSION Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 VERSION Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 KEYWORDS Cercopithecinae; Cercocebus.  
 SOURCE 1 (bases 1 to 668)  
 ORGANISM Villinger, F., Bostik, P., Mayne, A., King, C.L., Genain, C.P.,  
 Weiss, W.R. and Ansari, A.A.  
 TITLE Cloning, sequencing, and homology analysis of nonhuman primate  
 Pas/Fas-ligand and co-stimulatory molecules  
 JOURNAL Immunogenetics 53 (4), 315-328 (2001)  
 MEDLINE 21383618  
 PUBMED 11491535  
 REFERENCE 2 (bases 1 to 668)  
 AUTHORS Villinger, F.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-FEB-2001) Pathology and Laboratory Medicine, Emory University School of Medicine, Winship Cancer Institute, 1365B Clifton Rd, Atlanta, GA 30322, USA  
 FEATURES Location/Qualifiers  
 source 1. .668  
 /organism="Cercocebus torquatus atys"  
 /mol\_type="mRNA"  
 /sub\_species="atys"  
 /db\_xref="taxon:9531"  
 CDS 2. .664  
 /codon\_start=1  
 /product="CD28 protein"  
 /protein\_id="AAK37601.1"  
 /db\_xref="GI:1365471"  
 /translation="MLRLLALNLLPSIRVTGNKILVKQSPMLVAYDNAVNLCKYSY  
 NLFPSRFRASLHKGLDSAVEVCVYGNYSQQLQYISKTGFCGCKLGDSVTFYLQNM  
 YNQTDIYCKIEVMYPPPLYDKNGKNTIIHVKGKHLCPSPFPGSGKPFWALVVVGN  
 GLVACYSLLVTTVAFSIFCMRSKRSLHSDYNNMTPRPGTRKHQPYAPPRDFAAY  
 RS"





REFERENCE 2 (bases 1 to 668)  
AUTHORS Vallingner, F.  
TITLE Direct Submission  
JOURNAL Submitted (02-FEB-2001) Pathology and Laboratory Medicine, Emory University School of Medicine, Winship Cancer Institute, 1365B Clifton Rd, Atlanta, GA 30322, USA

FEATURES  
source Location/Qualifiers  
1..668  
/organism="Macaca mulatta"  
/mol\_type="mRNA"  
/db\_xref="taxon:9544"  
2..664  
/note="co-stimulatory molecule"  
/codon\_start=1  
/product="CD28 protein precursor"  
/protein\_id="AAK37538.1"  
/db\_xref="GI:13650008"  
/translation="MRLLLALNLLPSIQVTGNKILVKQSPMLVAYDNVNLCKYSY  
NLFSREFRSLHKLGLSAVEVCVVYEDYSQOLQVSKTEFNCQDKLGNSVTFYLNLI  
YVNOTDIYCKIEVMYPPYLDNEKSNGLIIHKGKHLCPSPLPFGPSKPFVALVVG  
GVLCAYSLLVTVAFSIFWRSKRSLHSDYMNWTPRPPTKKHYQPCAPPDRFAAY  
RS"  
2..55  
/replace="t"  
102 /replace="g"  
142 /replace="t"  
210 /replace="g"  
212 /replace="a"  
249 /replace="g"  
365 /replace="a"  
524 /replace="t"  
547 /replace="g"  
627 /replace="a"

ORIGIN  
Query Match 68.0%; Score 20.4; DB 9; Length 668;  
Best Local Similarity 95.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCCAGGAGCGATAGGCTGC 30  
Db 668 CGTGTCCAGGAGCGATAGGCTGC 647

RESULT 8  
AR102241/c  
LOCUS AR102241 875 bp DNA linear PAT 14-FEB-2001  
DEFINITION Sequence 1 from patent US 6083751.  
ACCESSION AR102241  
VERSION AR102241.1 GI:12813039  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 875)  
AUTHORS Feldhaus, A. Lawrence, and Jones, L. Ann.  
TITLE Chimeric receptors for the generation of selectively-activatable TH-independent cytotoxic T cells  
JOURNAL Patent: US 6083751-A 1 04-JUL-2000;  
FEATURES  
source Location/Qualifiers  
1..875  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 68.0%; Score 20.4; DB 9; Length 668;  
Best Local Similarity 95.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCCAGGAGCGATAGGCTGC 30  
Db 668 CGTGTCCAGGAGCGATAGGCTGC 647

RESULT 8  
AR102241/c  
LOCUS AR102241 875 bp DNA linear PAT 14-FEB-2001  
DEFINITION Sequence 1 from patent US 6083751.  
ACCESSION AR102241  
VERSION AR102241.1 GI:12813039  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 875)  
AUTHORS Feldhaus, A. Lawrence, and Jones, L. Ann.  
TITLE Chimeric receptors for the generation of selectively-activatable TH-independent cytotoxic T cells  
JOURNAL Patent: US 6083751-A 1 04-JUL-2000;  
FEATURES  
source Location/Qualifiers  
1..875  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 68.0%; Score 20.4; DB 6; Length 875;  
Best Local Similarity 95.5%; Pred. No. 1.1e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCCAGGAGCGATAGGCTGC 30  
Db 868 CGTGTCCAGGAGCGATAGGCTGC 847

RESULT 9  
BD267803/c  
LOCUS BD267803 1514 bp DNA linear PAT 17-JUL-2003  
DEFINITION Methods and products related to metabolic interactions in disease.  
ACCESSION BD267803  
VERSION BD267803.1 GI:33077571  
KEYWORDS JP 2002512203-A/6.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1514)  
AUTHORS Newell, M. K.  
TITLE Methods and products related to metabolic interactions in disease  
JOURNAL Patent: JP 2002512203-A 6 23-APR-2002;  
COMMENT THE UNIVERSITY OF VERMONT AND STATE AGRICULTURAL COLLEGE  
OS Homo sapiens (human)  
PN JP 2002512203-A/6  
PD 23-APR-2002  
PP 30-MAR-1999 JP 2000544356  
PR 17-APR-1998 US 60/082250, 29-JUL-1998 US 60/094519 PR  
24-SEP-1998 US 60/101580  
PT MARTHA K NEWELL  
PC A61K39/395, A61K31/519, A61K31/55, A61K31/704, A61K31/7056, A61K31/7088.  
PC A61K38/17, A61K38/21, A61K45/06, A61P3/10, A61P35/00, C12N15/00, PC G01N33/574,  
PC A61K37/12, A61K37/66, C12N15/00  
CC Methods and products related to metabolic interactions in CC disease  
FH Key Location/Qualifiers  
FT source 1..1514  
FT /organism="Homo sapiens (human)"  
FEATURES  
source Location/Qualifiers  
1..1514  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 68.0%; Score 20.4; DB 6; Length 1514;  
Best Local Similarity 95.5%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCCAGGAGCGATAGGCTGC 30  
Db 766 CGTGTCCAGGAGCGATAGGCTGC 745

RESULT 10  
I07254/c  
LOCUS I07254 1514 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 11 from Patent EP 0330191.  
ACCESSION I07254  
VERSION I07254.1 GI:590064  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1514)  
AUTHORS Seed, B. D. M. B., Allen, J., Aruffo, A., Camerini, D., Lauffer, L. D., Oquendo, C. P., Simmons, D., Stamenkovic, I., and Stengelin, S. D.

```

TITLE      Rapid immunoselection cloning method
JOURNAL    Patent: EP 030191-A2 11 30-AUG-1989;
FEATURES   Location/Qualifiers
source     1..1514
           /organism="unknown"
           /mol_type="unassigned DNA"

ORIGIN
Query Match      68.0%; Score 20.4; DB 6; Length 1514;
Best Local Similarity 95.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  9  CTTGTCAGGAGCGATAGGCTGC 30
      |||||
Db   766  CGTGTCTCAGGAGCGATAGGCTGC 745

RESULT 11
AR380771/c
LOCUS      AR380771 1514 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 1316 from patent US 6607879.
ACCESSION  AR380771
VERSION     AR380771.1 GI:40088405
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1514)
AUTHORS   Cocks, B.G., Stuart, S.G. and Seilhamer, J.J.
TITLE     Compositions for the detection of blood cell and immunological
          response gene expression
JOURNAL    Patent: US 6607879-A 1316 19-AUG-2003;
FEATURES   Location/Qualifiers
source     1..1514
           /organism="unknown"
           /mol_type="genomic DNA"

ORIGIN
Query Match      68.0%; Score 20.4; DB 6; Length 1514;
Best Local Similarity 95.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  9  CTTGTCAGGAGCGATAGGCTGC 30
      |||||
Db   766  CGTGTCTCAGGAGCGATAGGCTGC 745

RESULT 12
BD015187/c
LOCUS      BD015187 1514 bp DNA linear PAT 27-AUG-2002
DEFINITION TL15a cell surface antigen and CD27 cell surface antigen, and
          utilization thereof.
ACCESSION  BD015187
VERSION     BD015187.1 GI:22555994
KEYWORDS   JP 2001157592-A/5.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1514)
AUTHORS   Seed, B., Aruffo, A. and Amiot, M.
TITLE     TL15a cell surface antigen and CD27 cell surface antigen, and
          utilization thereof
JOURNAL    Patent: JP 2001157592-A 5 12-JUN-2001;
          THE GENERAL HOSPITAL CORP
COMMENT    OS Homo sapiens (human)
          PN JP 2001157592-A/5
          PD 12-JUN-2001
          PF 04-OCT-2000 JP 2000305557
          PR 13-JUL-1990 US 553759
          PI BRIAN SEED, ALEJANDRO ARUFFO, MARTIN AMIOT
          PC C12N15/09, C07K14/725, G01N33/53, C12N15/00
          CC TL15a cell surface antigen and CD27 cell surface antigen, and

FEATURES   Location/Qualifiers
source     1..1514
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /map="2q33-q34"
           1..1514
           /gene="CD28"
           100..762
           /gene="CD28"
           /note="T-cell-specific homodimer surface protein
           precursor"
           /codon_start=1
           /protein_id="AAA60581.1"
           /db_xref="GI:338445"
           /translation="MLRLALNLPFSIQVTGNKILVKQSPMLVAYDNAVNLCKYSY
           NLFSREFRASHLHGLDQSAVEVCVYGVNYSQQLQVYSGTGCNCDGKLGNESVTFLQNL
           YNQTDIYFCKIEVMYPPPYLDNEKSNGLIIHVKGKHLCPSPLPFGPSKPPFWLVVVG
           GVLACSLVLTVAFIIFVWRKRSRLHSDYNNMTPRRPGTKKHQYQYAPPRDFAAY
           RS"
           100..153
           /gene="CD28"
           /note="T-cell-specific homodimer surface protein signal
           peptide"
           154..759
           /gene="CD28"
           /product="T-cell-specific homodimer surface protein"

ORIGIN
Unreported.

```



Matches	21;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	9	CTTGTCTCAGGAGCGATAGGCTGC	30						
Db	889	CGTGTCTCAGGAGCGATAGGCTGC	868						
RESULT 19									
CQ715869/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
FEATURES									
source									
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
Qy	9	CTTGTCTCAGGAGCGATAGGCTGC	30						
Db	889	CGTGTCTCAGGAGCGATAGGCTGC	868						
RESULT 20									
AX283559/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
FEATURES									
source									
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
Qy	9	CTTGTCTCAGGAGCGATAGGCTGC	30						
Db	889	CGTGTCTCAGGAGCGATAGGCTGC	868						
RESULT 21									
AX283559/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
FEATURES									
source									
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
Qy	9	CTTGTCTCAGGAGCGATAGGCTGC	30						
Db	889	CGTGTCTCAGGAGCGATAGGCTGC	868						



```
repeat_region 105260..105545
/rpt_family="Alu"
repeat_region 105891..106161
/rpt_family="Alu"

ORIGIN
Query Match 68.0%; Score 20.4; DB 9; Length 106539;
Best Local Similarity 95.5%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTGTCTCAGGAGCGATAGGCTGC 30
Db 69947 CGTGTCTCAGGAGCGATAGGCTGC 69926

RESULT 23
AC125238/c
LOCUS AC125238 196622 bp DNA linear PRI 15-OCT-2002
DEFINITION Homo sapiens BAC clone RP11-711C24 from 2, complete sequence.
ACCESSION AC125238
VERSION AC125238.5 GI:22367884
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Sulston, J.E. and Waterston, R.
JOURNAL Toward a complete human genome sequence
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 196622)
AUTHORS Nguyen, C., Bielicki, L. and Schatzkammer, K.
TITLE The sequence of Homo sapiens BAC clone RP11-711C24
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 196622)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 196622)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 196622)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 196622)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 16, 2002 this sequence version replaced gi:22138709.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0711C24

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.
```

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC069314 and AC010138.

A transposon was identified in the finished region of this clone and removed prior to submission. The transposon would insert after base pair 83916 of this sequence.

The sequence from base 156374 to base 156453 was derived from one plasmid subclone.

The sequence from base 156479 to base 156499 was derived from one plasmid subclone.

Polymorphisms have been identified between AC069314, AC010138 and this sequence.

Data from AC069314 and AC010138 was used to finish this clone.

#### FEATURES

##### source

1. .196622  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-711C24"  
/clone\_lib="RPCI-11"

215. .236  
/rpt\_family="L1"  
240\_520  
/rpt\_family="Alu"  
892. .1086  
/rpt\_family="L1"  
1575. .1885  
/rpt\_family="Alu"  
1991. .2078  
/rpt\_family="MIR"  
2079. .2397  
/rpt\_family="Alu"  
2418. .2487  
/rpt\_family="Alu"  
2488. .2764  
/rpt\_family="Alu"  
2771. .2871  
/rpt\_family="MIR"  
3123. .3190  
/rpt\_family="L2"  
3296. .3599  
/rpt\_family="Alu"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region 3612. .3919  
/rpt\_family="Alu"  
repeat\_region 3926. .3988  
/rpt\_family="MIR"  
repeat\_region 4192. .4493  
/rpt\_family="Alu"  
repeat\_region 5165. .5256  
/rpt\_family="ACHobo"  
5257. .5593  
/rpt\_family="L1"  
repeat\_region 5599. .5623  
/rpt\_family="(TG)n"  
5624. .5748  
/rpt\_family="Alu"  
5752. .5929  
/rpt\_family="L1"  
6043. .6552  
/rpt\_family="ACHobo"  
6670. .6803  
/rpt\_family="L2"  
6804. .7097  
/rpt\_family="Alu"  
7098. .7192  
/rpt\_family="L2"  
7193. .7515  
/rpt\_family="Alu"  
7516. .7745  
/rpt\_family="L2"  
7746. .8054  
/rpt\_family="Alu"  
8055. .8264  
/rpt\_family="L2"  
8265. .8565  
/rpt\_family="Alu"  
8566. .9242  
/rpt\_family="L2"  
9381. .9682  
/rpt\_family="Alu"  
9688. .9997  
/rpt\_family="Alu"  
10134. .10384  
/rpt\_family="L2"  
10773. .11075  
/rpt\_family="Alu"  
11076. .11379  
/rpt\_family="Alu"  
11492. .11580  
/rpt\_family="MIR"  
11927. .12020  
/rpt\_family="MIR"  
12146. .12237  
/rpt\_family="MER2\_type"  
12238. .12534  
/rpt\_family="Alu"  
12535. .12682  
/rpt\_family="MER2\_type"  
12683. .13170  
/rpt\_family="MaLR"  
13171. .13193  
/rpt\_family="MER2\_type"  
13621. .14034  
/rpt\_family="MaLR"  
14683. .14966  
/rpt\_family="L1"  
14967. .15250  
/rpt\_family="Alu"  
15251. .15303  
/rpt\_family="L1"  
15796. .16074  
/rpt\_family="Alu"  
16075. .16315

Query Match

68.0%; Score 20.4; DB 9; Length 196622;









```
repeat_region      site:EcoRI"
repeat_region      complement(2750..2965)
repeat_region      /rpt_family="Tigger6"
repeat_region      3059..3493
repeat_region      /rpt_family="L1MDa"
repeat_region      3510..3851
repeat_region      /rpt_family="MTE"
repeat_region      3987..4684
repeat_region      /rpt_family="L1MDa"
repeat_region      4778..4992
repeat_region      /rpt_family="L1MDa"
repeat_region      4996..5962
repeat_region      /rpt_family="L1_MM"
repeat_region      6019..6218
repeat_region      /rpt_family="L1MDa"
repeat_region      6493..6560
repeat_region      /rpt_family="AT_rich"
repeat_region      complement(6613..7564)
repeat_region      /rpt_family="MT-int"
repeat_region      complement(7579..7665)
repeat_region      /rpt_family="MTD"
repeat_region      7666..7685
repeat_region      /rpt_family="(T)n"
repeat_region      complement(7686..8207)
repeat_region      /rpt_family="L1_MM"
repeat_region      complement(8208..8349)
repeat_region      /rpt_family="MTC"
repeat_region      complement(8352..9904)
repeat_region      /rpt_family="L1_RN"
repeat_region      complement(9905..10147)
repeat_region      /rpt_family="MTC"
repeat_region      10217..10428
repeat_region      /rpt_family="L1MC/D"
repeat_region      11350..11390
repeat_region      /rpt_family="(TA)n"
repeat_region      12150..12711
repeat_region      /rpt_family="Lx2"
repeat_region      12712..12748
repeat_region      /rpt_family="(A)n"
repeat_region      12756..12858
repeat_region      /rpt_family="Lx2"
repeat_region      12832..12983
repeat_region      /rpt_family="Lx2"
repeat_region      complement(13001..13415)
repeat_region      /rpt_family="Lx"
repeat_region      13405..14051
repeat_region      /rpt_family="Lx2"
repeat_region      14062..14090
repeat_region      /rpt_family="(CAAAA)n"
repeat_region      14109..14207
repeat_region      /rpt_family="B2_Mm2"
repeat_region      complement(14513..14629)
repeat_region      /rpt_family="B3"
repeat_region      14838..14880
repeat_region      /rpt_family="(TTTA)n"
repeat_region      complement(14882..15657)
repeat_region      /rpt_family="Lx4"
repeat_region      15700..16118
repeat_region      /rpt_family="L1"
repeat_region      complement(16298..17814)
repeat_region      /rpt_family="Lx2"
repeat_region      complement(17820..17915)
repeat_region      /rpt_family="Lx2"
repeat_region      17933..18932
repeat_region      /rpt_family="Lx2"
repeat_region      18955..19295
repeat_region      /rpt_family="Lx5"
repeat_region      19785..19903
repeat_region      /rpt_family="MIR"
repeat_region      complement(20487..21667)
repeat_region      /rpt_family="L1_MM"
repeat_region      22750..22770
repeat_region      /rpt_family="AT_rich"

complement(23503..25414)
/rpt_family="L1_MM"
25401..25714
/rpt_family="L1"
25715..26058
/rpt_family="IAPEY_LTR"

Query Match      67.3%; Score 20.2; DB 10; Length 216521;
Best Local Similarity 88.0%; Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6      GATCTTGTCTCAGGAGCATAGGCTGC 30
          |||||
Db      89113 GAGCTTGTCTCAGGAGCATAGGCGGC 89089
          |||||

RESULT 27
AC113935/c      142422 bp      DNA      linear      PRI 22-MAY-2002
LOCUS      Homo sapiens chromosome 1 clone RP4-706A16, complete sequence.
DEFINITION      AC113935
ACCESSION      AC113935
VERSION      AC113935.2 GI:21070697
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 142422)
AUTHORS      Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
      Saenphimmachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
      Haugen, E.D.
      Direct Submission
      Unpublished
      2 (bases 1 to 142422)
AUTHORS      Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
      Direct Submission
      Submitted (05-MAR-2002) Genome Center, University of Washington,
      Box 352145, Seattle, WA 98195, USA
      3 (bases 1 to 142422)
AUTHORS      Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
      Saenphimmachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
      Haugen, E.D.
      Direct Submission
      Submitted (22-MAY-2002) Genome Center, University of Washington,
      Box 352145, Seattle, WA 98195, USA
      On May 22, 2002 this sequence version replaced gi:19111744.
      ----- Genome Center
      Center: University of Washington Genome Center
      Center Code: UWGC
      Web site: http://www.genome.washington.edu
      Contact: uwgchtgs@u.washington.edu
      ----- Project Information
      Center project name: chr-1
      Center Clone name: RP4-706A16 (sc0534)
      ----- Summary Statistics
      Sequencing vector: plasmid; 100% of reads
      Chemistry: Dye-terminator ET; 53% of reads
      Chemistry: Dye-terminator Big Dye; 47% of reads
      Assembly program: Phrap; version 0.990319
      Consensus quality: 142336 bases at least Q40
      Consensus quality: 142421 bases at least Q30
      Insert size: 142422; sum-of-contigs
      Quality coverage: 8.6x in Q20 bases; sum-of-contigs
      -----
      Overlapping Sequences:
      5': RP4-564M11 AL035409
      3': RP11-363H12 (UWGC:sc0369) AC093433
      -----
      Sequence Quality Assessment:
      This entry has been annotated with sequence quality
      estimates computed by the Phrap assembly program.
      All manually edited bases have been reduced to quality zero.
```

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

#### Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

#### HindIII

#### BglII

#### EcoRI

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
3338	3384	11257	11189	3005	3039
449	<800	5671	5735	2184	2227
512	<800	10817	11189	8065	8193
2814	2859	659	<800	486	<800
1247	1216	5351	5311	106	<800
4695	4677	3324	3320	4142	4160
3373	3384	1771	1862	184	<800
2620	2688	293	<800	1815	1788
1870	1913	851	856	3300	3251
1692	1683	15640	15580	43	<800
5193	5137	1876	1862	994	977
6417	6365	1363	1323	1101	1107
939	950	90	<800	593	<800
7404	7516	1206	1212	1574	1524
301	<800	1438	1425	3233	3251
1932	1913	4283	4295	6425	6475
5602	5541	2805	2882	489	<800
6755	6842	749	760	3031	3039
2221	2177	509	<800	960	977
588	<800	5803	5735	10765	10646
2611	2688	11496	11189	4250	4160

#### FEATURES

source

#### Location/Qualifiers

1. 142422  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP4-706A16"  
/clone\_lib="RPCI human PAC library 4"

Query Match	66.7%	Score 20;	DB 9;	Length 142422;
Best Local Similarity	82.1%	Pred. No. 3e+02;		
Matches	23;	Conservative	0;	Mismatches 5; Indels 0; Gaps 0;
QY	1	TCGAGGATCTTGTGCAGAGCGATAGGCT 28		
DB	111649	TTGTGGATCTTGTGCAGAGTGATGGCT 111622		
RESULT 28				
AC098295/c				
LOCUS	AC098295	158498 bp	DNA	linear
DEFINITION	Rattus norvegicus clone CH230-1113, *** SEQUENCING IN PROGRESS ***;			HTG 13-NOV-2002
ACCESSION	3 unorderd pieces.			
VERSION	AC098295.4	GI:24941567		
KEYWORDS	HTG; HTGS; PHAS1; HTGS DRAFT; HTGS_ENRICHED.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
AUTHORS	1 (bases 1 to 158498)			
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Almbrooks,S.L., Amarathunga,H.C., Are,J.R., Ayele,M., Banks,T., Barabara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douchevaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,X., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheeshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S., Ogum,M., Okwuon,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Renvy,G., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savay,G., Scher,S., Scott,G., Shen,H., Shoohtant,N., Sisson,I., Sodergren,E., Sotaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,B., Vinson,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.			
TITLE	Direct Submission			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 158498)			
AUTHORS	Worley,K.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
REFERENCE	3 (bases 1 to 158498)			
AUTHORS	Worley,K.C.			

**COMMENT**

DEFINITION Homo sapiens chromosome 1 clone RP11-363H12, complete sequence.  
 ACCESSION AC093433 AL390781  
 VERSION AC093433.2 GI:21954025  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 172613)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.  
 and Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 172613)  
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and  
 Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-AUG-2001) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 REFERENCE 3 (bases 1 to 172613)  
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.  
 and Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-JUL-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 COMMENT On Jul 24, 2002 this sequence version replaced gi:15281284.  
 ----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UWGC  
 Web site: <http://www.genome.washington.edu>  
 Contact: [uwgchgs@u.washington.edu](mailto:uwgchgs@u.washington.edu)  
 Drafting Center: SC  
 ----- Project Information  
 Center project name: chr-1  
 Center clone name: RP11-363H12 (sc0369)  
 ----- Summary Statistics  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator ET; 89% of reads  
 Chemistry: Dye-terminator Big Dye; 11% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 172300 bases at least Q40  
 Consensus quality: 172573 bases at least Q30  
 Consensus quality: 172605 bases at least Q20  
 Insert size: 172613; sum-of-contigs  
 Quality coverage: 10.0x in Q20 bases; sum-of-contigs  
 -----  
 Overlapping Sequences:  
 5': RP4-706A16 (UWGC:sc0534) AC113935, 49862-bp overlap  
 3': RP11-375A5 (UWGC:sc0306) AC095030, 1917-bp overlap  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 Genbank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----  
 This sequence was finished as follows unless otherwise noted:  
 all regions were either double-stranded or sequenced with an  
 alternate chemistry or covered by high quality data (i.e., Phred  
 quality >= 30); an attempt was made to resolve all sequencing  
 problems, such as compressions and repeats; all regions were  
 covered by at least one plasmid subclone or more than one M13  
 subclone; and the assembly was confirmed by restriction digest.  
 -----  
 Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest  
 fingerprinting. Comparison of the experimentally derived digest

fragments with sequence-predicted fragments is given below.  
 The electronically-digested sequence consists of both insert and  
 vector, in order to accurately represent the entire circular BAC.  
 Small fragments below a variable cutoff (approximately 400-800 bp)  
 are not resolved in the fingerprint and hence do not appear  
 in the table. There are no significant remaining discrepancies  
 between the experimental and predicted values. Uniquely ordered  
 fragments are separated by dashed lines.

BgIII				EcoRI				HindIII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2745	2737	8863	10108	10313							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2067	2064	6	6382	6310							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9628	9798	841	512	<800							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9759	9798	1070	449	<800							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
862	859	14180	3030	3039							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
6337	6355	716	5750	5702							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
327	<800	234	10459	10313							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
4183	4152	3847	2026	2031							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
971	994	1973	980	993							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3551	3535	2483	1772	1761							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1007	994	117	2062	2031							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9815	9798	1021	10	<800							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
663	<800	4538	9101	8985							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3886	3851	2518	11909	11985							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3878	3851	3451	1585	1550							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1112	1100	11127	1619	1550							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2227	2208	8028	4791	4569							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1600	1547	2736	1197	1170							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3771	3851	3887	2447	2464							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7721	7711	6016	4288	4250							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5015	5005	64	611	<800							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
546	<800	1118	640	<800							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
13578	13550	2243	11334	11289							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
550	<800	213	472	<800							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2797	2842	1486	1309	1311							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3309	3288	953	4116	4050							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1477	1441	2403	2430	2464							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3532	3535	4552	3061	3039							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5347	5354	876	5163	5114							
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7658	7711	808	1532	1550							

FEATURES		Location/Qualifiers		1. .172613	
source					
Query Match		66.7%; Score 20; DB 9; Length 172613;			
Best Local Similarity		82.1%; Pred. No. 3.le+02;			
Matches 23; Conservative		0; Mismatches 5; Indels 0; Gaps 0;			
QY	1	TCGAGGATCTTGTCAGGAGCGATAGGCT	28		
Db	19078	TTGTGGATCTTGTCAGGAGTGATGGCT	19051		
RESULT 31					
AC073027/c		173995 bp DNA linear HTG 25-JUN-2000			
LOCUS					
DEFINITION		Homo sapiens chromosome 1 clone RP11-295E21 map 1, WORKING DRAFT			
ACCESSION					
VERSION		AC073027.2 GI:8705069			
KEYWORDS		HTG; HTGS PHASE1; HTGS_DRAFT.			
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens			

REFERENCE	6310
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	3039
AUTHORS	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 173995)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 1, clone RP11-295E21  
Unpublished  
2 (bases 1 to 173995)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karataas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Liu,C., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Melgrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (08-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 25, 2000 this sequence version replaced gi:8347970.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L10424  
Center clone name: 295\_E\_21  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 159136 bases at least Q40  
Consensus quality: 167456 bases at least Q30  
Consensus quality: 170233 bases at least Q20  
Insert size: 171595; sum-of-contigs  
Quality coverage: 3.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1530: contig of 1530 bp in length  
\* 1531 1630: gap of 100 bp  
\* 1631 3100: contig of 1470 bp in length  
\* 3101 3200: gap of 100 bp  
\* 3201 4696: contig of 1496 bp in length  
\* 4697 4797: gap of 100 bp  
\* 4797 6486: contig of 1690 bp in length  
\* 6487 6587: gap of 100 bp  
\* 6587 8765: contig of 2179 bp in length  
\* 8766 8866: gap of 100 bp  
\* 8866 11068: contig of 2203 bp in length  
\* 11069 11168: gap of 100 bp



## ORIGIN

**Qy**      1 TCGAGGATCTTGTACGAGCGTAGGCT 28  
          | | | | | | | | | | | | | |  
**Db**      111862 TTGTGGATCTTGTGAGAGTGATTGGCT 111835

RESULT 32  
AC026874/c  
LOCUS

AC026874/C	AC026874	203812 bp	DNA	linear	HTG 01-SEP-2000
LOCUS	Homo sapiens chromosome 1 clone RP11-627E4, WORKING DRAFT SEQUENCE,				
DEFINITION	32 unordered pieces.				
ACCESSION	AC026874.4				
VERSION	GI:9958302				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 203812)				
JOURNAL	Waterston,R.H.				
REFERENCE	The sequence of Homo sapiens clone				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 203812)				
JOURNAL	Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (24-MAR-2000) Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
JOURNAL	MO 63108, USA				
COMMENT	On Sep 1, 2000 this sequence version replaced qi:7631093.				

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH062YE04
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 186139 bases at least Q40

```

```
/clone="RP11-295E21"
/clone.lib="RP11-11 Human M-10 P10"
```

[illegible]

1.	1530	/note="assembly_fragment"
	1631.	13100
	3201.	4695
	4797.	6486
	6587.	8765
	8866.	11068
	11169.	13797
	13898.	16098
	16199.	18174
	18275.	20802
	20903.	23111
	23212.	25268
	25369.	27582
	27683.	31473

Consensus quality: 190766 bases at least Q30  
Consensus quality: 192884 bases at least Q20  
Insert size: 192000; agarose-bp  
Insert size: 200712; sum-of-contigs  
Quality coverage: 3.83 in Q20 bases; agarose-bp  
Quality coverage: 3.78 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2006: contig of 2006 bp in length  
\* 2007: gap of unknown length  
\* 2107: contig of 2682 bp in length  
\* 4788: gap of unknown length  
\* 4888: gap of unknown length  
\* 7839: contig of 2951 bp in length  
\* 7840: gap of unknown length  
\* 7940: 11627: contig of 3688 bp in length  
\* 11628: gap of unknown length  
\* 11728: contig of 4391 bp in length  
\* 16118: gap of unknown length  
\* 16219: contig of 3461 bp in length  
\* 19679: gap of unknown length  
\* 19680: contig of 3901 bp in length  
\* 23681: gap of unknown length  
\* 23781: contig of 5139 bp in length  
\* 28920: gap of unknown length  
\* 33880: contig of 4861 bp in length  
\* 33981: gap of unknown length  
\* 39616: contig of 5636 bp in length  
\* 39717: gap of unknown length  
\* 39717: 46646: contig of 6930 bp in length  
\* 46647: gap of unknown length  
\* 46747: 52566: contig of 5820 bp in length  
\* 52567: gap of unknown length  
\* 52667: contig of 6652 bp in length  
\* 59318: gap of unknown length  
\* 59419: contig of 6498 bp in length  
\* 65917: gap of unknown length  
\* 66017: contig of 6302 bp in length  
\* 72319: gap of unknown length  
\* 72419: contig of 8360 bp in length  
\* 80779: gap of unknown length  
\* 80879: contig of 6035 bp in length  
\* 86914: gap of unknown length  
\* 87014: 94398: contig of 7385 bp in length  
\* 94399: gap of unknown length  
\* 94499: 104139: contig of 9641 bp in length  
\* 104140: 104239: gap of unknown length  
\* 104240: 113608: contig of 9369 bp in length  
\* 113609: gap of unknown length  
\* 113709: 114866: contig of 1158 bp in length  
\* 114867: 114967: gap of unknown length  
\* 114967: 128022: contig of 13056 bp in length  
\* 128023: 128123: gap of unknown length  
\* 139097: contig of 10975 bp in length  
\* 139098: gap of unknown length  
\* 139198: 149544: contig of 10347 bp in length  
\* 149545: gap of unknown length  
\* 149645: 162226: contig of 12582 bp in length  
\* 162227: 162326: gap of unknown length  
\* 162327: 178058: contig of 15732 bp in length  
\* 178059: gap of unknown length  
\* 178159: 195531: contig of 17373 bp in length  
\* 195532: 195632: gap of unknown length  
\* 195633: 196744: contig of 1113 bp in length  
\* 196745: 196844: gap of unknown length  
\* 196845: 198567: contig of 1723 bp in length  
\* 198568: 198667: gap of unknown length

\* 198668 200798: contig of 2131 bp in length  
\* 200799 200898: gap of unknown length  
\* 200899 202307: contig of 1409 bp in length  
\* 202308 202407: gap of unknown length  
\* 202408 203812: contig of 1405 bp in length.  
FEATURES  
Location/Qualifiers  
source  
1. .203812  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-627E4"  
misc\_feature  
1. .2006  
/note="assembly\_name:Contig10"  
misc\_feature  
2107. .4788  
/note="assembly\_name:Contig11"  
misc\_feature  
4889. .7839  
/note="assembly\_name:Contig12"  
misc\_feature  
7940. .11627  
/note="assembly\_name:Contig13"  
misc\_feature  
11728. .16118  
/note="assembly\_name:Contig14"  
misc\_feature  
16219. .19679  
/note="assembly\_name:Contig15"  
misc\_feature  
19780. .23680  
/note="assembly\_name:Contig16"  
misc\_feature  
23781. .28919  
/note="assembly\_name:Contig17"  
misc\_feature  
29020. .33880  
/note="assembly\_name:Contig18"  
misc\_feature  
33981. .39616  
/note="assembly\_name:Contig19"  
misc\_feature  
39717. .46646  
/note="assembly\_name:Contig20"  
misc\_feature  
46747. .52566  
/note="assembly\_name:Contig21"  
misc\_feature  
52667. .59318  
/note="assembly\_name:Contig22"  
misc\_feature  
59419. .65916  
/note="assembly\_name:Contig23"  
misc\_feature  
66017. .72318  
/note="assembly\_name:Contig24"  
misc\_feature  
72419. .80778  
/note="assembly\_name:Contig27"  
misc\_feature  
80879. .86913  
/note="assembly\_name:Contig25"  
misc\_feature  
87014. .94398  
/note="assembly\_name:Contig26"  
misc\_feature  
94499. .104139  
/note="assembly\_name:Contig28"  
misc\_feature  
104240. .113608  
/note="assembly\_name:Contig29"  
misc\_feature  
113709. .114866  
/note="assembly\_name:Contig3"  
misc\_feature  
114967. .128022  
/note="assembly\_name:Contig30"  
misc\_feature  
128123. .139097  
/note="assembly\_name:Contig31"  
misc\_feature  
139198. .149544  
/note="assembly\_name:Contig32"  
misc\_feature  
149645. .162226  
/note="assembly\_name:Contig33"  
misc\_feature  
162327. .178058  
/note="assembly\_name:Contig34"  
misc\_feature  
178159. .195531  
/note="assembly\_name:Contig35"  
misc\_feature  
195632. .196744  
/note="assembly\_name:Contig4"  
misc\_feature  
196845. .198567  
/note="assembly\_name:Contig5"  
misc\_feature  
198668. .200798  
/note="assembly\_name:Contig7"  
200899. .202307

/note="assembly\_name:Contig9"  
202408..203812  
/note="assembly\_name:Contig9"

ORIGIN

Query Match 66.7%; Score 20; DB 2; Length 203812;  
Best Local Similarity 82.1%; Pred.No. 3.2e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TCGAGGATCTTGTCAGGACGGATAGGCT 28  
||||| ||||| ||||| ||||| |||||  
Db 86013 TTGTGGATCTTGTGAGGAGTATTGGCT 85986

RESULT 33  
AC115240

LOCUS  
DEFINITION Rattus norvegicus clone CH230-153F4, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC115240  
ACCESSION AC115240  
VERSION AC115240.6 GI:25139449  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS Muzny,D., Marie,J., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,B., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blythe,P., Brown,M., Bryant,N., Buhatz,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,R., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Angelis,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huliyil,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L.S., Jiang,H., Johnson,B., Johnson,R., Jollivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Loulueghed,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,B., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newcom,N., Nguyen,N., Norris,S., Nwakoleleh,O., Okwuonu,G., Olarnpunsaagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfamkotch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajda,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taborski,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejo,Z., Usmani,K., Valdes,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,

```

end_sequence:BH336181"
misc_feature
228890..230652
/note="wgs_contig"
misc_feature
230703..232735
/note="wgs_contig"
ORIGIN
Query Match 66.7%; Score 20; DB 2; Length 232735;
Best Local Similarity 82.1%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CGAGATCTTCTCAGGACGATAGCTG 29
Db 218720 CTAGATCTTCTCAGGACGATAGCTG 218747

RESULT 34
AC120701
LOCUS AC120701 237445 bp DNA linear HTG 21-SEP-2002
DEFINITION Rattus norvegicus clone CH230-65H6, *** SEQUENCING IN PROGRESS ***,
4 unordered pieces.
ACCESSION AC120701
VERSION AC120701.4 GI:23265381
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 237445)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blewalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devita,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gabregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,D., Jacob,D., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseg,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Manguney,S., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokeleneh,O., Okunnu,G., Olarnpungsoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puaio,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Rielly,B., Rielly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojao,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soza,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabar,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,

```

Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 237445)  
 Worley,K.C.  
 Direct Submission  
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 237445)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (21-SEP-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Sep 21, 2002 this sequence version replaced gi:21908396.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
 sequence may extend beyond the ends of the clone and there may be  
 contigs that consist entirely of whole genome shotgun sequence  
 reads. Both end sequences and whole genome shotgun sequence only  
 contigs will be indicated in the feature table.  
 ----- Genome Center -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information -----  
 Center project name: GXQV  
 Center clone name: CH230-65H6  
 ----- Summary Statistics -----  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 209781 bases at least Q40  
 Consensus quality: 213033 bases at least Q30  
 Consensus quality: 214997 bases at least Q20  
 Estimated insert size: 233017; sum-of-contigs estimation  
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 233866: contig of 233866 bp in length  
 \* 233867 233866: gap of unknown length  
 \* 233967 235011: contig of 1045 bp in length  
 \* 235012 235111: gap of unknown length  
 \* 235112 236137: contig of 1026 bp in length  
 \* 236138 236237: gap of unknown length  
 \* 236238 237445: contig of 1208 bp in length.  
 \* Location/Qualifiers  
 1. 237445  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-65H6"  
 misc\_feature  
 1..1326  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 8065..8944  
 /note="clone\_boundary  
 clone\_end:T7"  
 misc\_feature  
 FEATURES  
 source

```

misc_feature
  site: ECoRI
  end_sequence: BH350813"
  complement(232953..233569)
  /note="clone boundary
  clone end: Sp6
  site: ECoRI
  end_sequence: BH350815"

ORIGIN
Query Match      66.7%; Score 20; DB 2; Length 237445;
Best Local Similarity 82.1%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 GAGGATCTTCTCAGGAGCGATGAGCTGTC 30
    |||||
DB 23224 GAGGATCGTGTCTCAGCAGCGCTGAGGCTGC 23251

RESULT 35
AC129059
DEFINITION
Rattus norvegicus clone CH230-124M19, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION
AC129059.4 GI:30580087
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 243439)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Deigado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Loreshuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,S., Norris,S.,
Nwaokelimeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Popovic,D., Primus,E., Pu,L.,
Plopper,F., Poindexter,A., Popovic,D., Reeves,K., Regier,M.A., Reigh,R.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Richards,S., Riggs,F.,
Reilly,B., Reilly,J., Ren,Y., Reuter,M., Rose,R., Ruiz,S.J.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,

```

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V.,  
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 243439)  
Worley, K.C.  
Direct Submission  
Submitted (25-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 243439)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:23264545.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GLIE  
Center clone name: CH230-124M19  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 238140 bases at least Q40  
Consensus quality: 239753 bases at least Q30  
Consensus quality: 240841 bases at least Q20  
Estimated insert size: 250221: sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 2 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
\* 1 242268: contig of 242268 bp in length  
\* 242269 242368: gap of unknown length  
\* 242369 243439: contig of 1071 bp in length.

FEATURES  
source :  
1. 243439  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-124M19"  
misc\_feature 347..1113  
/notes="clone\_boundary"

```

clone_end:T7
site:ECORI
end_sequence:BH329260"
complement(240371..241219)
/notes="clone boundary
clone_end:Sp6
site:ECORI
end_sequence:BH329263"

ORIGIN
Query Match          66.7%; Score 20; DB 2; Length 243439;
Best Local Similarity 82.1%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GAGGATCTTGTTCAGGAGCGATAGGCTGC 30
||||| ||||| ||||| ||||| |||||
Db 237133 GAGGATCGTGTTCAGCAGCGCTGAGGCTGC 237160

RESULT 36
AC127373/c
LOCUS          280288 bp      DNA      linear      HTG 25-AUG-2002
DEFINITION    Mus musculus chromosome UNK clone RP23-42M8, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
AC127373
AC127373.1 GI:21747787
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
VERSION       Mus musculus
KEYWORDS      Mus musculus (house mouse)
SOURCE        Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 280288)
              McPherson,J.D. and Waterston,R.H.
              The sequence of Mus musculus clone
              Unpublished
REFERENCE     2 (bases 1 to 280288)
              McPherson,J.D. and Waterston,R.H.
              Direct Submission
              Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
REFERENCE     3 (bases 1 to 280288)
              McPherson,J.D. and Waterston,R.H.
              Direct Submission
              Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M BA0042M08
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator; 100%
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 276868 bases at least Q40
Consensus quality: 277312 bases at least Q30
Consensus quality: 277488 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
1 4562: contig of 4562 bp in length

```

```

* 4563 4662: gap of unknown length
* 4563 37815: contig of 33153 bp in length
* 37816 37915: gap of unknown length
* 37916 66758: contig of 28843 bp in length
* 66759 66858: gap of unknown length
* 66859 99532: contig of 32674 bp in length
* 99533 99632: gap of unknown length
* 99633 141409: contig of 41777 bp in length
* 141410 141509: gap of unknown length
* 141510 200444: contig of 58935 bp in length
* 200445 200544: gap of unknown length
* 200545 270671: contig of 70127 bp in length
* 270672 270771: gap of unknown length
* 270772 274472: contig of 3701 bp in length
* 274473 274573: gap of unknown length
* 274573 276805: contig of 2233 bp in length
* 276806 276906: gap of unknown length
* 276906 280288: contig of 3383 bp in length.

```

## FEATURES

```

source
1..280288
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-42M8"
misc_feature
1..4562
/notes="assembly_name:Contig10"
misc_feature
4563..37815
/notes="assembly_name:Contig11"
misc_feature
37916..66758
/notes="assembly_name:Contig12"
misc_feature
66859..99532
/notes="assembly_name:Contig13"
misc_feature
99633..141409
/notes="assembly_name:Contig14"
misc_feature
141510..200444
/notes="assembly_name:Contig15"
misc_feature
200545..270671
/notes="assembly_name:Contig16"
misc_feature
270772..274472
/notes="assembly_name:Contig17"
misc_feature
274573..276805
/notes="assembly_name:Contig18"
misc_feature
276906..280288
/notes="assembly_name:Contig19"

```

## ORIGIN

```

Query Match          66.7%; Score 20; DB 2; Length 280288;
Best Local Similarity 82.1%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 3 GAGGATCTTGTTCAGGAGCGATAGGCTGC 30
||||| ||||| ||||| ||||| |||||

```

```

Db 170584 GAGCATCTTGCCATGAGCGATATGCAGC 170557
||||| ||||| ||||| ||||| |||||

```

```

RESULT 37
AP004596
LOCUS

```

```

AP004596 300050 bp      DNA      linear      BCT 17-MAY-2003
DEFINITION    Oceanobacillus iheyensis HTE831 genomic DNA, section 4/13.
ACCESSION     AP004596 BA000028
VERSION       AP004596.1 GI:22776527
KEYWORDS
SOURCE        Oceanobacillus iheyensis HTE831
ORGANISM      Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.

```

```

REFERENCE
AUTHORS
TITLE

```

```

Lu,J., Nogi,Y. and Takami,H.
Oceanobacillus iheyensis gen. nov., sp. nov., a deep-sea extremely
halotolerant and alkaliphilic species isolated from a depth of 1050
m on the Iheya Ridge
JOURNAL       FEMS Microbiol. Lett. 205 (2), 291-297 (2001)
MEDLINE       21623015

```

PUBMED 11750818  
 REFERENCE 2  
 AUTHORS Takami,H., Takaki,Y. and Uchiyama,I.  
 TITLE Genome sequence of Oceanobacillus iheyensis isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments  
 JOURNAL Nucleic Acids Res. 30 (18), 3927-3935 (2002)  
 MEDLINE 2220767  
 PUBMED 12235376  
 REFERENCE 3 (bases 1 to 300050)  
 AUTHORS Takami,H., Takaki,Y. and Chee,G.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-DEC-2001) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group; 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan  
 (E-mail:takami@jamstec.go.jp,  
 URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,  
 Tel:81-468-67-9643, Fax:81-468-67-9645)

FEATURES  
 source  
 1..300050  
 /organism="Oceanobacillus iheyensis HTE831"  
 /mol\_type="genomic DNA"  
 /strain="HTE831"  
 /db\_xref="taxon:221109"  
 /note="halotolerant and alkaliphilic"  
 154..441  
 /gene="OB0847"  
 154..441  
 /gene="OB0847"  
 /note="CDS ID OB0847"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical conserved protein"  
 /protein\_id="BAC12803.1"  
 /db\_xref="GI:22776528"  
 /translation="MATISKGENKFVVGEDIRNPOAEITFVSEGENRLVIDHTYVATD  
 LRQGGIAQQLLDHVITYAREKGYVPLCPFAKGOIQKQKYNVLNVQHI"  
 466..696  
 /gene="OB0848"  
 466..696  
 /gene="OB0848"  
 /note="CDS ID OB0848"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical conserved protein"  
 /protein\_id="BAC12804.1"  
 /db\_xref="GI:22776529"  
 /translation="MKVITPEAAEYLSMSESTVSALVLOGKINAIHDGKQYLVNKKQOF  
 THFEQVEKMKMIEYQLQEPEDLDVQDED"  
 775..1686  
 /gene="OB0849"  
 775..1686  
 /gene="OB0849"  
 /note="CDS ID OB0849"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical conserved protein"  
 /protein\_id="BAC12805.1"  
 /db\_xref="GI:22776530"  
 /translation="MKKRPPRSKKEPASITYEVEEQILLPFLKVMNRSNVS  
 ILTRGVSDHDIETKHNFLNKGSTVRILKNKAAIQDDVFGMSIYEDDDIIVIEK  
 ESSLGSIATYERKERTALHQLMHYVQKQHPANRVYVHRLDKDTSGVMFAKNRYKR  
 KLQNNKWDIVKERTYVALVEQVAKGKIVTSWLKESATRKMYSSRVQNDGQPAATRY  
 RVLQTNRYSLLEQLQETGRKNGIRVHMEELGHPVVDGKRYGSKGRSVIGRLGLHAKI  
 LSPYHVTKMLFRSDVNPAPQTK"  
 complement (1804..2790)  
 /gene="OB0850"  
 complement (1804..2790)  
 /gene="OB0850"  
 /note="CDS ID OB0850"  
 /codon\_start=1  
 /transl\_table=11  
 /product="D-aminopeptidase"

gene  
 CDS  
 /protein\_id="BAC12806.1"  
 /db\_xref="GI:22776531"  
 /translation="MAKSGDKDCTDVRGVQVGHVTLVEKINEKDTICTGVTAILPH  
 PGNLFRQVACSVINGFGKNTGLVQLNELGIESPIMLTNTLSIAPVWQGLTYML  
 NRETEIGDTGTINLVINGCNSYLNLSIQALITPSHAIKAIASNSSESSEAGVAG  
 TQWQCFYGGIGTASRLIQVEDVIYTLGGLVSVNYGRKQATFAWEHTGQETPDGS  
 IMWILATDPLSDROLNRIAKCAAGLGRGTSALDHGSGDIAIAFATANQVSHDSKNH  
 LMSAAYLRDDHPVWNLFOAVVEVIESCVIHSIQKAKTILGRSGHKVKAPL"  
 complement (2863..3459)  
 /gene="OB0851"  
 complement (2863..3459)  
 /gene="OB0851"  
 /note="CDS ID OB0851"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC12807.1"  
 /db\_xref="GI:22776532"  
 /translation="MKKWKVKPTIILGTVLILFGFLIFVDOVEKETNATIKNSDEN  
 EKEVVEKDPLESEETKENANPFGDMVSASGMNRRLIKYMHFMSHOKVEAQSKWGFY  
 EITDRIAPLQESFNTDSNAFEVKERASIRNILDRLWSNGDFSQVSDHNTLYRNSHE  
 DEDRGMATILSEDQEKFNKPFENKDFSKIRGINRFDDE"  
 3616..4107  
 /gene="OB0852"  
 3616..4107  
 /gene="OB0852"  
 /note="CDS ID OB0852"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome c biogenesis (thiol:disulfide  
 interchange protein)"  
 /protein\_id="BAC12808.1"  
 /db\_xref="GI:22776533"  
 /translation="MKILKYVLIIISIMFLPIMTAYAEGTVEGERAPDFELKTDGQO  
 LRLSPFKGRVLINFTWCPQCRQMPQMFYQDLQNLAVNLTDENKQVVR  
 QSLEQLTPIILLDEKGEVSKAYRISPIPTTWMIDSEGIIRHKSYGALTYEQMVAEYN  
 KME"  
 4237..4470  
 /gene="OB0853"  
 4237..4470  
 /gene="OB0853"  
 /note="CDS ID OB0853"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical conserved protein"  
 /protein\_id="BAC12809.1"  
 /db\_xref="GI:22776534"  
 /translation="WDKGTIIRTIVLALTNQFLITAGVHPITPGTQELWGEILSSIF  
 TIVATLTAFKNVYTYKKGKQHVLEHQLAK"  
 4496..5290  
 /gene="OB0854"  
 4496..5290  
 /gene="OB0854"  
 /note="CDS ID OB0854"  
 /codon\_start=1  
 /transl\_table=11  
 /product="N-acetylmuramoyl-L-alanine amidase (sporulation  
 mother cell wall hydrolase)"  
 /protein\_id="BAC12810.1"  
 /db\_xref="GI:22776535"  
 /translation="MKRIFIDPGHGNDFGAMNGIQLKDKINDIALRMKEYLLQTPG  
 GHEVLRSDRTKTVLSLQRTTWANNQADLYSIHNAGGTFESTYFNGIYSGKAE  
 THRRLQVQHSIVNETGYDRGMKEANFHLRESQMPAVLTENGFDYREDANKLKS  
 SFLOSIAKSHATGADVLELDPCEGASQGVYELADSLWTYNSPEWEDKAVIVHKGSEV  
 FTVIKDKFPVGGHMYQIKSGLYITANKVYRYTSTSTSKVKSVLN"  
 5378..5578  
 /gene="OB0855"  
 5378..5578  
 /gene="OB0855"  
 /note="CDS ID OB0855"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"

protein\_id="BAC12811.1"  
/db\_xref="GI:22776536"  
/translation="MKLLVQLVITALLWGMASFYDQMNELNTWIFYLVTSMWLLLL  
VLVKEWFLRKSTKEDKINE"  
5593. 7476  
/genes="OB0856"  
5593. 7476  
/genes="OB0856"  
/notes="CDS ID OB0856"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter ATP-binding protein"  
/protein\_id="BAC12812.1"  
/db\_xref="GI:22776537"  
/translations="MLMIDNLSKTYGDKTLFNOITATIGEKERIGLGVNGTKSKSFL  
KVTAGMSPDQGIKSKDVHIEVYQETNLOEDLTVLQIYVGDALIMRVMEQVBOA  
LLDLENSPEDAALQKLIQMQQKMDVEAEANTAKTLTKLGINTFTTVEKLSGG  
QKRIALAKALIOPADLIIMDEPTNHLHDHTVLEWLEKYSQSYSGVLLVTHDRFLNR  
VTNRIFELDKGNLYTAGNYVLEFLEKKAERLESNEQKHQNTKRELAWLRGAKA  
RSTKQARVREEMKDKTFTDKKEVFTQAGSKRLGNDVIEIDGLYKSGFNQVLSN  
FSQIIPEDRIGIIGNNGTKTLLNMAKRTPEPDQGYVKIGETVLIGYVTOGEELD  
DOORIIEYIKDVAEVIHTKGEVITAEOMLERFLPDAQQWSFIRTLGGERRLVLL  
KVLKSPNVLFLEPNTDLDLQSLVLEDEYLEHPFGVITVSHDRFLDRVYNKLIVF  
KNGNTDTFGNTELEKQAEKLOPTPKQKSSSTKISQNEKKKKLSMKEGWDTI  
EDEITKLEBEIVEIEQAIVDAGSDSEKQVQELYANQQAVALAEKRWAEALSIRKVEQ  
LET"  
7769. 8821  
/genes="OB0857"  
7769. 8821  
/genes="OB0857"  
/notes="CDS ID OB0857"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical conserved protein"  
/protein\_id="BAC12813.1"  
/db\_xref="GI:22776538"  
/translations="MQEIVSDVITFRGSHYQFGKMQGSEIKRSFVNMENKKQWKLRRP  
RFTINIDEVMSMINTYAPGIVWELEGIKDGIENWLSDTLQEFAGYINLPSSCSYFT  
SSNYMIRNVDYHPKTYEGRVLVFPDPTDGLTIGPSQRITGRLDGMNQYGLVGNFT  
Query Match 66.7%; Score 20; DB 1; Length 300050;  
Best Local Similarity 82.1%; Pred. No. 3.3e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCT 28  
Db 71007 TCATGTTCTTGTTCAGGAGCGATAGGCT 71034  
RESULT 38  
BX005159/c 312728 bp DNA linear HTG 08-JAN-2003  
LOCUS Homo sapiens chromosome 1 clone RP11-619A13.  
DEFINITION BX005159  
ACCESSION BX005159.2 GI:27650315  
VERSION HTG; HTGS PHASE2; HTGS\_CANCELLED.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Hall,R.  
Direct Submission  
Submitted (07-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
On Jan 11, 2003 this sequence version replaced gi:27368635.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information

Center project name: BA619A13  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 167037 bases at least Q40  
Consensus quality: 167127 bases at least Q30  
Consensus quality: 167156 bases at least Q20  
Insert size: 312728; sum-of-contigs  
Insert size: 161110; 3.9% error; agarose-fp  
Quality coverage: 4.1lx in Q20 bases; sum-of-contigs Quality  
coverage: 8.0lx in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 312728: contig of 312728 bp in length.  
FEATURES  
source  
1. 312728  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-619A13"  
/clone\_lib="RPC1-11.3"  
1. 312728  
/note="assembly\_fragment:01914  
clone\_end:SP6  
vector\_side:left  
clone\_end:T7  
vector\_side:right"  
misc\_feature  
1. 312728  
/note="assembly\_fragment:01914  
clone\_end:SP6  
vector\_side:left  
clone\_end:T7  
vector\_side:right"  
ORIGIN  
Query Match 66.7%; Score 20; DB 2; Length 312728;  
Best Local Similarity 82.1%; Pred. No. 3.3e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCT 28  
Db 19078 TTGTGGATCTTGTTCAGGAGCGATAGGCT 19051  
RESULT 39  
AC136270/c 331221 bp DNA linear HTG 23-NOV-2002  
LOCUS Rattus norvegicus clone CH230-46911, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 12 unordered pieces.  
AC136270 AC136270.2 GI:25188310  
VERSION HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 331221)  
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Diya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,



Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geert, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hume, J., Hulyk, S., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakiemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Prannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivas, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

# TITLE JOURNAL

REFERENCE  
2 (bases 1 to 331221)

AUTHORS  
Rat Genome Sequencing Consortium.

TITLE  
Direct Submission

Submitted (31-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 331221)

AUTHORS  
Rat Genome Sequencing Consortium.

TITLE  
Direct Submission

Submitted (23-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 23, 2002 this sequence version replaced gi:24431541. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: KBUS

Center clone name: CH230-46911

Summary Statistics

Assembly program: Phrap; version 0.950329  
Consensus quality: 257314 bases at least Q40  
Consensus quality: 265389 bases at least Q30  
Consensus quality: 271049 bases at least Q20  
Estimated insert size: 267830; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html)  
NOTE: This sequence may represent more than one clone.  
NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 4872: contig of 4872 bp in length  
4873 4972: gap of unknown length  
4973 25330: contig of 20358 bp in length  
25331 25430: gap of unknown length  
25431 112631: contig of 87201 bp in length  
112632 112731: gap of unknown length  
112732 262442: contig of 149711 bp in length  
262443 262442: gap of unknown length  
262443 306822: contig of 44280 bp in length  
306823 317614: contig of 10692 bp in length  
317615 317714: gap of unknown length  
317715 319314: contig of 1600 bp in length  
319315 319414: gap of unknown length  
319415 320449: contig of 1035 bp in length  
320450 320550: gap of unknown length  
320551 322112: contig of 1463 bp in length  
322113 324044: contig of 1932 bp in length  
324045 324145: gap of unknown length  
324146 328550: contig of 4406 bp in length  
328551 331221: contig of unknown length  
331222 331221: contig of 2571 bp in length.

## FEATURES source

1. 331221  
/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-46911"

25431..26656

/note="wgs contig"

77244..78677

/note="wgs contig"

112732..114588

/note="wgs contig"

152455..154397

/note="wgs contig"

158417..160948

/note="wgs contig"

172688..174086

/note="wgs contig"

202542..204106

/note="wgs contig"

244136..245292

/note="wgs contig"

262543..263907

/note="wgs contig"

## ORIGIN

Query Match 66.7%; Score 20; DB 2; Length 331221;  
Best Local Similarity 82.1%; Pred. No. 3.3e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2 CGAGGATCTGTGACGAGCGATAGGCTG 29

||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db      197084 CTAGATTCTCTCAAGAGGATAGGCTG 197057

RESULT 40
AY057108/c
LOCUS   5359 bp      DNA      linear      PLN 24-APR-2002
DEFINITION Hypocrea jecorina QM9414 transcriptional adaptor-like protein
            (ada2), galactose-1-phosphate uridylyltransferase (gal7), and
            transmembrane domain-containing protein (erg28) genes, complete
            cds.
ACCESSION AY057108
VERSION   AY057108.2 GI:18101631
KEYWORDS
SOURCE    Hypocrea jecorina
           Hypocrea jecorina
           Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
           Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 5359)
AUTHORS   Seiboth,B., Hofmann,G. and Kubicek,C.P.
TITLE      Lactose metabolism and cellulase production in Hypocrea jecorina:
            the gal7 gene, encoding galactose-1-phosphate uridylyltransferase,
            is essential for growth on galactose but not for cellulase
            induction
JOURNAL   Mol. Genet. Genomics 267 (1), 124-132 (2002)
MEDLINE   21918477
PubMed    11919723
REFERENCE 2 (bases 1 to 5359)
AUTHORS   Seiboth,B., Hofmann,G. and Kubicek,C.P.
TITLE      Direct Submission
JOURNAL   Submitted (25-SEP-2001) IBTM, TU Vienna, Getreidemarkt 9, Vienna
            1060, Austria
REFERENCE 3 (bases 1 to 5359)
AUTHORS   Seiboth,B., Hofmann,G. and Kubicek,C.P.
TITLE      Direct Submission
JOURNAL   Submitted (09-JAN-2002) IBTM, TU Vienna, Getreidemarkt 9, Vienna
            1060, Austria
REMARK    Sequence update by submitter
COMMENT   On Jan 9, 2002 this sequence version replaced gi:16209627.
FEATURES             source
     source            1..5359
                        /organism="Hypocrea jecorina"
                        /mol_type="genomic DNA"
                        /strain="QM9414"
                        /db_xref="taxon:51453"
                        complement(<361..>2015)
                        /gene="ada2"
                        complement(join(<361..1812,1923..>2015))
                        /gene="ada2"
                        /product="transcriptional adaptor-like protein"
                        complement(join(361..1812,1923..2015))
                        /gene="ada2"
                        /note="Ada2"
                        /codon_start=1
                        /product="transcriptional adaptor-like protein"
                        /protein_id="AAL14200.2"
                        /db_xref="GI:18101632"
                        /translation="MGVIRKKTATRGEGVKYVCVDCSDITSTVIRICADPACSDF
                        DLCPFCGSGESNAHDPAHFAVIEQNSFFIFDREWGAEELLLLEGAIYGLGSW
                        ADIADHIGFREDVDRHYLTYSIESPNFPLPERGCPNNELANEI PREEFOAKKR
                        RIEREASNAALQPKTKPTASVPSCHIEIQMPGRLEFETEYANEAEAVOLMOP
                        DPGDGINPKTGELEPEMELKLTWMEIYNCELQORVERKXVIFEHNLDDYRENTKQEKR
                        KTYEERDLQARFPFARIWNHDKDFEDFTQGLQDELNROIAQLQEWRSNKICDLRSQ
                        KYEAEAKARIQAIIPGSMRDERLASSQRSAQPPPPSPGASLLVAPELPIRSAQA
                        KNGANGEDKALNGVNGHSPAKRYQPIISGVPPLOLTQENADYHLLTPPEE
                        AKLCVLRILQPKFYLMIKEQIIKEAIKNGTLLKKQAKEICRLDSOKGARLFDFFSNA
                        GWVGKA"
                        <3150..>4505
                        /gene="gal7"
                        join(<3150..3250,3325..3376,3444..3579,3646..>4505)
                        /gene="gal7"
                        /product="galactose-1-phosphate uridylyltransferase"
                        join(3150..3250,3325..3376,3444..3579,3646..4505)
                        /gene="gal7"

     gene
     mRNA
     CDS

```

```

/EC number="2.7.7.12"
/note="Gal7"
/codon_start=1
/product="galactose-1-phosphate uridylyltransferase"
/protein_id="AAL14201.1"
/db_xref="GI:16209629"
/translation="MPDKILDDISHRRYNPLTDSWLLVSPHRTKRPWQGOEGAAVTT
LPEYDPKCYLCPGNSRAAGQPNPQYOTFAFVNDYSAVKEQQDPDYVDQSSDDLESLL
LRAQGVKGVYVLTSPSKINVTLADMSAKDILPTINHWTLYANHLSPNPLSAVAQ
LQPLISKEAPVPKONYRMYQIFENKGAAGGNSPHPHCQAWTISTMPPEPGKELVQM
AKYRQOGRHLLADIYKGLAKEERVVQNDSSFVVCVPMWAIWPFVEVLVLPKRVKRAL
VDLTADERLQLAIEAQEYTRRYDNLFECHFPYSSGHIQAPLDGTPEEINAYPHMEFY
PPLRSATVKYKFLVGFELMAEAQRDITPEQATIRLRACDGLYRNKLS"
complement(<4623..>5208)
/gene="erg28"
complement(join(<4623..5072,5146..>5208))
/gene="erg28"
/product="transmembrane domain-containing protein"
complement(join(4623..5072,5146..5208))
/gene="erg28"
/note="Erg28; similar to the Erg28p of Saccharomyces
cerevisiae"
/codon_start=1
/product="transmembrane domain-containing protein"
/protein_id="AAL14202.2"
/db_xref="GI:18101633"
/translation="MDLKSILPPAKGILPYMLILSVVSGNSLQAYATLHFSRRVY
NGFRIRNPKLPPASAKFNEDSPNKLVAQNDPKATDQLTPLAGRLFGTWTLITCIVR
CYAAHYHLHLGPVYNIAITWYVVALGHFASELFVFKSWTFGLQIPIFFTALTALIWP
LVRSHYVEFE"

ORIGIN
Query Match      65.3%; Score 19.6; DB 8; Length 5359;
Best Local Similarity 84.6%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TCGAGGATCTTGTTCAGGACGATAGG 26
        |||||
DB      3169 TCGAGGATCTTGTTCAGGACGATAGG 3144

Search completed: January 7, 2005, 11:33:35
Job time : 1039.21 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 930.789 Seconds  
(without alignments)  
1371.762 Million cell updates/sec

Title: US-09-786-502A-7  
Perfect score: 27  
Sequence: 1 gggcgcaattgaagtattatcct 27

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Capext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sta.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	27	6	BD272284
2	19.2	71.1	189912	2	AC102434
3	19.2	71.1	214580	10	AC102369
4	19.2	71.1	245862	2	AC123598
5	19	70.4	542	9	HUMCD282
6	19	70.4	663	6	AX695883
7	19	70.4	705	9	HSAS17504
8	19	70.4	1514	6	BD267803
9	19	70.4	1514	6	I07254
10	19	70.4	1514	6	AR380771
11	19	70.4	1514	6	BD015187
12	19	70.4	1514	9	HUMSPTCS
13	19	70.4	3803	6	AX557214
14	19	70.4	3804	6	AX695882
15	19	70.4	3806	6	CQ715869
16	19	70.4	3806	6	AX283559
17	19	70.4	51365	6	AX695881
18	19	70.4	66887	9	AF411057
19	19	70.4	88624	2	AC020676

C	20	19	70.4	88624	2	AC020676	AC020676 Homo sapi
	21	19	70.4	106539	9	AF225899	AF225899 Homo sapi
	22	19	70.4	108100	9	AC103881	AC103881 Homo sapi
	23	19	70.4	148587	2	CR407559	CR407559 Danio rer
	24	19	70.4	175531	9	HSXKSRPXM	AL121577 Homo sapi
	25	19	70.4	196622	9	AC125238	AC125238 Homo sapi
	26	19	70.4	235811	2	AC094790	AC094790 Rattus no
	27	18.6	68.9	151552	5	AL954182	AL954182 Zebrafish
C	27	18.6	68.9	200339	10	AC122833	AC122833 Mus muscu
C	28	18.6	68.9	218328	2	AC121459	AC121459 Rattus no
C	29	18.6	68.9	227835	10	AC141426	AC141426 Mus muscu
C	30	18.6	68.9	227835	10	AC141426	AC141426 Mus muscu
C	31	18.6	68.9	227929	2	AC094794	AC094794 Rattus no
C	32	18.6	68.9	236612	2	AC109678	AC109678 Rattus no
C	33	18.6	68.9	247100	2	AC100967	AC100967 Mus muscu
C	34	18.6	68.9	290066	2	AC129421	AC129421 Rattus no
C	35	18.4	68.1	146945	9	AC009464	AC009464 Homo sapi
C	36	18.4	68.1	180254	2	AC110187	AC110187 Mus muscu
C	37	18.4	68.1	193005	10	AC110917	AC110917 Mus muscu
C	38	18.4	68.1	199432	2	AC107233	AC107233 Mus muscu
C	39	18.4	68.1	242301	2	AC010172	AC010172 Homo sapi
C	40	18.2	67.4	1465	8	SOE02	X79823 S.occidenta
C	41	18.2	67.4	1639	5	BS935170	BS935170 Gallus ga
C	42	18.2	67.4	3615	9	HSN803200	AL831875 Homo sapi
C	43	18.2	67.4	3694	6	BD246112	BD246112 Developme
C	44	18.2	67.4	56506	6	BD246111	BD246111 Developme
C	45	18.2	67.4	87302	9	AC027313	AC027313 Homo sapi

ALIGNMENTS

RESULT 1  
BD272284  
LOCUS  
DEFINITION Fused receptor specific to prostate-specific membrane antigen and  
its use.  
ACCESSION BD272284  
VERSION BD272284.1 GI:33082052  
KEYWORDS JP 2002524081-A/7.  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 27)  
AUTHORS Sadelain,M., Bander,N.H. and Gong,M.  
TITLE Fused receptor specific to prostate-specific membrane antigen and  
its use  
JOURNAL Patent: JP 2002524081-A/7 06-AUG-2002;  
SLOAN KETTERING INSTITUTE FOR CANCER RESEARCH  
COMMENT OS Synthetic construct  
PN JP 2002524081-A/7  
PD 06-AUG-2002  
PF 03-SEP-1999 JP 2000568998  
PI MICHEL SADELAIN,NEIL H BANDER,MICHAEL GONG  
PC C07K19/00;  
PC C12N15/09,A61K35/26,A61K39/395,A61K48/00,A61P35/00,  
CC C12N5/10//C07K16/44,C12N15/00,C12N5/00  
CC Fused receptor specific to prostate-specific membrane antigen  
and its use  
FH Key Location/Qualifiers  
FT source 1..27  
FT Location/Qualifiers  
source 1..27  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

Query Match 100.0%; Score 27; DB 6; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.011; 0;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCCGAATTGAAGTTATGATCCT 27  
 |||||  
 Db 1 GCGCGCCGAATTGAAGTTATGATCCT 27

## RESULT 2

AC102434/c

LOCUS

DEFINITION

AC102434

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC102434 189912 bp DNA linear HTG 25-MAR-2004  
 Mus musculus chromosome 12 clone RP24-120L6 map 12, \*\*\* SEQUENCING  
 IN PROGRESS \*\*\*, 5 unordered pieces.

AC102434 4 GI:45736689  
 HTG; HTGS PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
 Mus musculus (house mouse)

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 189912)

Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus chromosome 12, clone RP24-120L6

Unpublished  
 2 (bases 1 to 189912)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kells, C., Lakocque, K.,  
 Lamarez, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,  
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 189912)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V.,  
 Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J.,  
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
 DeArelano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L.,  
 Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,  
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagoopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
 MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,  
 McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 189912)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V.,  
 Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J.,  
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,  
 DeArelano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L.,  
 Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D.,  
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagoopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,  
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,  
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,  
 MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,  
 McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,  
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS



```

repeat_region      /rpt family="Lx6"
                    complement(17621..17854)
repeat_region      /rpt family="Li_MW"
                    17950..17993
repeat_region      /rpt family="AT_rich"
                    18088..18162
repeat_region      /rpt family="AT_rich"
                    18181..18210
repeat_region      /rpt family="AT_rich"
                    18344..18374
repeat_region      /rpt family="AT_rich"
                    18929..19138
repeat_region      /rpt family="RLTR10C"
                    19139..19166
repeat_region      /rpt family="(GA)n"
                    19167..19364
repeat_region      /rpt family="RLTR10C"
                    complement(20397..20500)
repeat_region      /rpt family="nigger7"
                    complement(20569..20859)
repeat_region      /rpt family="Tigger7"
                    complement(20938..21158)
repeat_region      /rpt family="B4A"
                    21360..21382
repeat_region      /rpt family="AT_rich"
                    complement(21486..21826)
repeat_region      /rpt family="Lx8"
                    21849..21891
repeat_region      /rpt family="(TG)n"
                    complement(21904..22282)
repeat_region      /rpt family="ORR1B2"
                    complement(22283..22499)

Query Match      71.1%; Score 19.2; DB 10; Length 214580;
Best Local Similarity 87.5%; Pred. NO. 2.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3  GGCGCAATTCAGCTATGTATCC 26
Db 8449 GGCTCAATTCAGATGTATCC 8472

RESULT 4
AC123598
LOCUS      AC123598.4 GI:42761736
DEFINITION Mus musculus chromosome 12 clone RP24-501E5 map 12, *** SEQUENCING
IN PROGRESS ***, 7 unordered pieces.
ACCESSION  AC123598
VERSION     AC123598.4
KEYWORDS   HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE   1. (bases 1 to 245862)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeKrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., Fitzgerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Levine,R.,
            Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neill,D., Oliver,J., Peterson,K., Lander,E., Abouelleil,A., Allen,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 245862)
AUTHORS    Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (24-FEB-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 24, 2004 this sequence version replaced gi:28570471.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L17504
            Center clone name: 501_E_5
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 19304: contig of 19304 bp in length
            * 19305 19404: gap of 100 bp
            * 19405 41155: contig of 21751 bp in length
            * 41156 41255: gap of 100 bp
            * 41256 155335: contig of 114080 bp in length
            * 155336 155435: gap of 100 bp
            * 155436 165910: contig of 10475 bp in length
            * 165911 166010: gap of 100 bp
            * 166011 184480: contig of 18470 bp in length
            * 184481 184581: gap of 100 bp
            * 184581 212061: contig of 27461 bp in length
            * 212062 212161: gap of 100 bp
            * 212162 245862: contig of 33701 bp in length.
            *
            * Location/Qualifiers
            * 1. 245862
            * /organism="Mus musculus"
            * /mol_type="genomic DNA"
FEATURES
source

```

```

/db_xref="taxon:10090"
/chromosome="12"
/map="12"
/clone="RP24-501E5"
/clone_lib="RPCI-24 Male Mouse BAC"

ORIGIN

Query Match      71.1%; Score 19.2; DB 2; Length 245862;
Best Local Similarity 87.5%; Pred. No. 2.e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy   3  GGCGCAATTGAAGTTATGTCCT 26
    ||| ||||| ||| ||||| |||
Db   204043  GGCTCAATGCAGATATGTCCT 204066

RESULT 5
LOCUS HUMCD282          542 bp     DNA       linear      PRI 01-NOV-1994
DEFINITION Human T-cell membrane glycoprotein CD28 mRNA, exon 2.
ACCESSION M37813
VERSION M37813.1 GI:180087
KEYWORDS glycoprotein CD28.
SEGMENT 2 of 4
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS Lee,K.P., Taylor,C., Petryniak,B., Turka,L.A., June,C.H. and Thompson,C.B.
TITLE The genomic organization of the CD28 gene. Implications for the regulation of CD28 mRNA expression-and heterogeneity
JOURNAL J. Immunol. 145 (1), 344-352 (1990)
MEDLINE 90293482
PUBMED 2162892
COMMENT Original source text: Human DNA.
FEATURES             Location/Qualifiers
     source           1..542
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="2q33-q34"
                     order(M37813.1:1813..984,1..64)
                     /gene="CD28"
                     /note="G00-118-765"
                     /number=1
     intron           65..421
                     /gene="CD28"
                     /note="G00-118-765"
                     /number=2

ORIGIN

Query Match      70.4%; Score 19; DB 9; Length 542;
Best Local Similarity 100.0%; Pred. No. 2.e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   351  AATTGAAGTTATGTCCT 27
    ||| ||||| ||| ||||| |||
Db   351  AATTGAAGTTATGTCCT 369

RESULT 6
LOCUS AX695883          663 bp     DNA       linear      PAT 31-MAR-2003
DEFINITION Sequence 1510 from Patent WO03008583.
ACCESSION AX695883
VERSION AX695883.1 GI:29419048
KEYWORDS 
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```





```
/db_xref="taxon:9606"

ORIGIN
Query Match 70.4%; Score 19; DB 6; Length 1514;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
Db 438 AATTGAAGTTATGATCCT 456

RESULT 12
HUMSPCTS 1514 bp mRNA linear PRI 13-JAN-1995
LOCUS Human T-cell-specific homodimer surface protein CD28 mRNA, complete
DEFINITION cds.
ACCESSION J02988
VERSION J02988.1 GI:338444
KEYWORDS T-cell-specific homodimer surface protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1514)
Aruffo, A. and Seed, B.
Molecular cloning of a CD28 cDNA by a high-efficiency COS cell
expression system
Proc. Natl. Acad. Sci. U.S.A. 84 (23), 8573-8577 (1987)
MEDLINE 88068631
PUBMED 2825196
COMMENT Original source text: Human T-cell tumor line HPB-ALL, cDNA to
mRNA, clone lambda-H3W.
Draft entry and computer-readable sequence for [1] kindly provided
by B. Seed, 11-AUG-1987.

FEATURES
source Location/Qualifiers
1..1514
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="2q33-q34"
1..1514
/genes="CD28"
100..762
/notes="T-cell-specific homodimer surface protein
precursor"
/codon_start=1
/protein_id="AA60581.1"
/db_xref="GI:338445"
/db_xref="GDB:G00-118-765"
/translations="MLRLILALNLPISQVTGNKILVKQSPMLVAYDNVNLCKYSY
NLPFRFRASLHKGLDSAVEVCVYGVYGNYSQQLQVYSGTGCDCGLGNSVTFYLQNL
VYNQTDLYCKIEWYPPPYLDNEKSNGLIIHVKGKHLCPSPFPQSPKPFVWLVVVG
GLVACSLVLTAVFIIFVWRKSRKRLHSDYMNTPRPGRKKHYQYAPPRDFAAY
RS"
sig_peptide 100..153
/genes="CD28"
/notes="T-cell-specific homodimer surface protein signal
peptide"
mat_peptide 154..759
/genes="CD28"
/product="T-cell-specific homodimer surface protein"
ORIGIN Unreported.

Query Match 70.4%; Score 19; DB 9; Length 1514;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
Db 438 AATTGAAGTTATGATCCT 456
```

```
RESULT 13
AX557214 3803 bp DNA linear PAT 27-NOV-2002
LOCUS Sequence 1 from Patent WO02066059.
DEFINITION AX557214
ACCESSION AX557214.1 GI:25900195
VERSION AX557214.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
O'Hara, R.M. and Nagelin, A.M.
Agents that specifically block cd28-mediated signaling and uses
therefor
Patent: WO 02066059-A 1 29-AUG-2002;
Genetics Institute, LLC (US)
FEATURES
source Location/Qualifiers
1..3803
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 70.4%; Score 19; DB 6; Length 3803;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
Db 560 AATTGAAGTTATGATCCT 578

RESULT 14
AX695882 3804 bp DNA linear PAT 31-MAR-2003
LOCUS Sequence 1509 from Patent WO03008583.
DEFINITION AX695882
ACCESSION AX695882
VERSION AX695882.1 GI:29419047
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
Morris, D.W. and Engelhard, E.K.
Novel compositions and methods for cancer
Patent: WO 03008583-A 1509 30-JAN-2003;
Sagres Discovery (US)
FEATURES
source Location/Qualifiers
1..3804
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 70.4%; Score 19; DB 6; Length 3804;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
Db 561 AATTGAAGTTATGATCCT 579

RESULT 15
CQ715869 3806 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 1803 from Patent WO02068579.
DEFINITION CQ715869
ACCESSION CQ715869
VERSION CQ715869.1 GI:42276726
KEYWORDS
```

```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1
AUTHORS     Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE       Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
JOURNAL     Patent: WO 02068579-A 1803 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES    source
            1..3806
            /location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      70.4%; Score 19; DB 6; Length 3806;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 AATTGAAGTTATGATCCT 27
        |||||
Db      561 AATTGAAGTTATGATCCT 579

RESULT 16
AX283559
LOCUS      AX283559          3806 bp          DNA          linear          PAT 20-NOV-2001
DEFINITION Sequence 3 from Patent WO0179300.
ACCESSION  AX283559
VERSION     AX283559.1 GI:17044297
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1
AUTHORS     Bluestone, J.A., Collins, M., Whitters, M., Griffin, M. and Kranz, D.
TITLE       Surface-bound antigen binding portions of antibodies that bind to
            ctla-4 and cd28 and uses therefor
JOURNAL     Patent: WO 0179300-A 3 25-OCT-2001;
            GENETICS INSTITUTE, INC. (US)
FEATURES    source
            1..3806
            /location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      70.4%; Score 19; DB 6; Length 3806;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 AATTGAAGTTATGATCCT 27
        |||||
Db      561 AATTGAAGTTATGATCCT 579

RESULT 17
AX695881
LOCUS      AX695881          51365 bp          DNA          linear          PAT 31-MAR-2003
DEFINITION Sequence 1508 from Patent WO03008583.
ACCESSION  AX695881
VERSION     AX695881.1 GI:29419046
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1
AUTHORS     Morris, D.W. and Engelhard, E.K.

```

---

```

TITLE       Novel compositions and methods for cancer
JOURNAL     Patent: WO 03008583-A 1508 30-JAN-2003;
            Sagres Discovery (US)
FEATURES    source
            1..51365
            /location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      70.4%; Score 19; DB 6; Length 51365;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 AATTGAAGTTATGATCCT 27
        |||||
Db      30447 AATTGAAGTTATGATCCT 30465

RESULT 18
AF411057
LOCUS      AF411057          66887 bp          DNA          linear          PRI 13-DEC-2001
DEFINITION Homo sapiens clone BAC_22606 CD28 antigen (CD28) gene, exons 1, 2
            and 3, and partial cds.
ACCESSION  AF411057
VERSION     AF411057.1 GI:17646225
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 66887)
AUTHORS     Ling, V., Wu, P.W., Finnerty, H.F., Agostino, M.J., Graham, J.R.,
            Chen, S., Jussiff, J.M., Fisk, G.J., Miller, C.P. and Collins, M.
            Assembly and Annotation of Human Chromosome 2q33 Sequence
            Containing the CD28, CTLA4, and ICOS Gene Cluster: Analysis by
            Computational, Comparative, and Microarray Approaches
            Genomics 78 (3), 155-168 (2001)
            11735222

TITLE       JOURNAL
PUBMED      11735222
REFERENCE   2 (bases 1 to 66887)
AUTHORS     Ling, V., Wu, P.W., Finnerty, H.F., Agostino, M.J., Graham, J.R.,
            Chen, S., Jussiff, J., Fisk, G.J., Miller, C.P. and Collins, M.
            Direct Submission
            Submitted (15-AUG-2001) Department of Immunology, Genetics
            Institute/Wyeth Research, 200 Cambridge Park Drive, Cambridge, MA
            02140, USA

FEATURES    source
            1..66887
            /location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2q33"
            /clone="BAC_22606"
            7838..8329
            /note="similar to NADH:ubiquinone oxidoreductase"
            42348..>66887
            /gene="CD28"
            /number=1
            mRNA      join(42348..42621,62505..62861,65540..>65664)
            /gene="CD28"
            /product="CD28 antigen"
            /number=1
            exon      42348..42621
            /gene="CD28"
            /number=1
            5'UTR      42348..42569
            /gene="CD28"
            CDS        join(42570..42621,62505..62861,65540..>65664)
            /gene="CD28"
            /codon_start=1
            /product="CD28 antigen"
            /protein_id="AAL40931.1"
            /db_xref="GI:17646226"
            /translation="MLRLLALNLFPSIQVTGNKILVKQSPMLVAYDNAVNLSCKYSY

```

NLFSRFRASLHKGLDSAVECVWYGNYSQOLQVYKTCFNCDCGLGNSVTFYQLNL  
 YNQDITVYFCKLEVMPPFLDNERKSNGLIIHVKGKHLCPFLFPGPSKPFVWLVVVG  
 GVLACSLVTVAFIIFW"

exon 62505..62861  
 /gene="CD28"  
 /number=2  
 exon 65540..65664  
 /gene="CD28"  
 /number=3

## ORIGIN

Query Match 70.4%; Score 19; DB 9; Length 66887;  
 Best Local Similarity 100.0%; Pred. No. 3e+02; 0; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0

QY 9 AATTGAAGTTATGTCCT 27

Db 62791 AATTGAAGTTATGTCCT 62809

## RESULT 19

AC020676 88624 bp DNA linear HTG 13-JUL-2000  
 LOCUS Homo sapiens clone RP11-8P23, LOW-PASS SEQUENCE SAMPLING.

AC020676

AC020676

AC020676.2 GI:9123792

HTG; HTGS\_PHASE0.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 88624)

Birren,B., Linton,L., Nusbaum,C. and Lander,B.

Homo sapiens, clone RP11-8P23

Unpublished

2 (bases 1 to 88624)

Birren,B., Linton,L., Nusbaum,C., Lander,B., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

Boguslavskiy,L., Boukigalter,B., Brown,A., Burkett,G., Castle,A.,

Choepel,Y., Collangelo,M., Collins,S., Collamore,A., Cooke,P.,

DeArelano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,

Gardyna,S., Grant,G., Hegos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Oliviar,T.M., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A. and Zody,M.

Direct Submission

Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6682672.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2995

Center clone name: 8\_P\_23

-----

\* NOTE: This record contains 88 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 \* 882: contig of 882 bp in length  
 \* 883  
 \* 882: gap of 100 bp  
 \* 886: contig of 904 bp in length  
 \* 1887  
 \* 1886: gap of 100 bp  
 \* 1987  
 \* 1986: contig of 919 bp in length  
 \* 2906  
 \* 2905: gap of 100 bp  
 \* 3006  
 \* 3005: contig of 898 bp in length  
 \* 3904  
 \* 4004: gap of 100 bp  
 \* 4925: contig of 922 bp in length  
 \* 4926  
 \* 5025: gap of 100 bp  
 \* 5926  
 \* 5924: contig of 899 bp in length  
 \* 5925  
 \* 6024: gap of 100 bp  
 \* 6025  
 \* 6377: contig of 913 bp in length  
 \* 6938  
 \* 7037: gap of 100 bp  
 \* 7038  
 \* 7949: contig of 912 bp in length  
 \* 7950  
 \* 8049: gap of 100 bp  
 \* 8921  
 \* 9020: contig of 871 bp in length  
 \* 9894: contig of 874 bp in length  
 \* 9895  
 \* 9994: gap of 100 bp  
 \* 9995  
 \* 10908: contig of 914 bp in length  
 \* 10909  
 \* 11921: contig of 913 bp in length  
 \* 11922  
 \* 12021: gap of 100 bp  
 \* 12022  
 \* 12931: contig of 910 bp in length  
 \* 12932  
 \* 13031: gap of 100 bp  
 \* 13032  
 \* 13947: contig of 916 bp in length  
 \* 13948  
 \* 14047: gap of 100 bp  
 \* 14048  
 \* 14949: contig of 902 bp in length  
 \* 14950  
 \* 15049: gap of 100 bp  
 \* 15050  
 \* 15960: contig of 911 bp in length  
 \* 15961  
 \* 16060: gap of 100 bp  
 \* 16061  
 \* 16971: contig of 911 bp in length  
 \* 16972  
 \* 17071: gap of 100 bp  
 \* 17072  
 \* 17996: contig of 925 bp in length  
 \* 17997  
 \* 18096: gap of 100 bp  
 \* 18097  
 \* 18986: contig of 890 bp in length  
 \* 18987  
 \* 19086: gap of 100 bp  
 \* 19087  
 \* 19992: contig of 906 bp in length  
 \* 19993  
 \* 20992: gap of 100 bp  
 \* 20993  
 \* 20958: contig of 866 bp in length  
 \* 20959  
 \* 21058: gap of 100 bp  
 \* 21059  
 \* 21948: contig of 890 bp in length  
 \* 21949  
 \* 22048: gap of 100 bp  
 \* 22049  
 \* 22957: contig of 909 bp in length  
 \* 22958  
 \* 23057: gap of 100 bp  
 \* 23058  
 \* 23967: contig of 910 bp in length  
 \* 23968  
 \* 24067: gap of 100 bp  
 \* 24068  
 \* 25001: contig of 934 bp in length  
 \* 25002  
 \* 25101: gap of 100 bp  
 \* 25102  
 \* 26000: contig of 899 bp in length  
 \* 26001  
 \* 26100: gap of 100 bp  
 \* 26101  
 \* 27013: contig of 913 bp in length  
 \* 27014  
 \* 27113: gap of 100 bp  
 \* 27114  
 \* 28031: contig of 918 bp in length  
 \* 28032  
 \* 28131: gap of 100 bp  
 \* 28132  
 \* 29041: contig of 910 bp in length  
 \* 29042  
 \* 29141: gap of 100 bp  
 \* 30038: contig of 897 bp in length  
 \* 30039  
 \* 30138: gap of 100 bp  
 \* 30139  
 \* 31031: contig of 893 bp in length  
 \* 31032  
 \* 31131: gap of 100 bp  
 \* 31132  
 \* 32049: contig of 918 bp in length  
 \* 32050  
 \* 32149: gap of 100 bp  
 \* 32150  
 \* 33064: contig of 915 bp in length  
 \* 33065  
 \* 33164: gap of 100 bp

\* 33165 34070: contig of 906 bp in length  
\* 34071 34170: gap of 100 bp  
\* 34171 35094: contig of 924 bp in length  
\* 35095 35194: gap of 100 bp  
\* 35195 36126: contig of 932 bp in length  
\* 36127 36226: gap of 100 bp  
\* 36227 37155: contig of 929 bp in length  
\* 37156 37255: gap of 100 bp  
\* 37256 38171: contig of 916 bp in length  
\* 38172 38271: gap of 100 bp  
\* 38272 39176: contig of 905 bp in length  
\* 39177 39276: gap of 100 bp  
\* 39277 40184: contig of 908 bp in length  
\* 40185 40284: gap of 100 bp  
\* 40285 41184: contig of 900 bp in length  
\* 41185 41284: gap of 100 bp  
\* 41285 42188: contig of 904 bp in length  
\* 42189 42288: gap of 100 bp  
\* 42289 43177: contig of 889 bp in length  
\* 43178 43277: gap of 100 bp  
\* 43278 44193: contig of 922 bp in length  
\* 44200 44299: gap of 100 bp  
\* 44300 45219: contig of 920 bp in length  
\* 45220 45319: gap of 100 bp  
\* 45320 46209: contig of 890 bp in length  
\* 46210 46309: gap of 100 bp  
\* 46310 47208: contig of 899 bp in length  
\* 47209 47308: gap of 100 bp  
\* 47309 48233: contig of 925 bp in length  
\* 48234 48333: gap of 100 bp  
\* 48334 49266: contig of 933 bp in length  
\* 49267 49366: gap of 100 bp  
\* 49367 50292: contig of 926 bp in length  
\* 50293 50392: gap of 100 bp  
\* 50393 51290: contig of 898 bp in length  
\* 51291 51390: gap of 100 bp  
\* 51391 52279: contig of 889 bp in length  
\* 52280 52379: gap of 100 bp  
\* 52379 53275: contig of 896 bp in length  
\* 53276 53375: gap of 100 bp  
\* 53376 54289: contig of 914 bp in length  
\* 54290 54389: gap of 100 bp  
\* 54390 55311: contig of 922 bp in length  
\* 55312 55411: gap of 100 bp  
\* 55412 56319: contig of 908 bp in length  
\* 56320 56419: gap of 100 bp  
\* 56420 57325: contig of 906 bp in length  
\* 57326 57425: gap of 100 bp  
\* 57426 58342: contig of 917 bp in length  
\* 58343 58442: gap of 100 bp  
\* 58443 59358: contig of 916 bp in length  
\* 59359 59458: gap of 100 bp  
\* 59459 60380: contig of 922 bp in length  
\* 60381 60480: gap of 100 bp  
\* 60481 61386: contig of 906 bp in length  
\* 61387 61486: gap of 100 bp  
\* 61487 62365: contig of 879 bp in length  
\* 62366 62465: gap of 100 bp  
\* 62466 63359: contig of 894 bp in length  
\* 63360 63459: gap of 100 bp  
\* 63460 64352: contig of 893 bp in length  
\* 64353 64452: gap of 100 bp  
\* 64453 65354: contig of 902 bp in length  
\* 65355 65454: gap of 100 bp  
\* 65455 66370: contig of 916 bp in length  
\* 66371 66470: gap of 100 bp  
\* 66471 67399: contig of 929 bp in length  
\* 67400 67499: gap of 100 bp  
\* 67500 68395: contig of 896 bp in length  
\* 68396 68495: gap of 100 bp  
\* 68496 69391: contig of 896 bp in length  
\* 69392 69491: gap of 100 bp  
\* 69492 70404: contig of 913 bp in length

\* 70405 70504: gap of 100 bp  
\* 70505 71416: contig of 912 bp in length  
\* 71417 71516: gap of 100 bp  
  
Query Match 70.4%; Score 19; DB 2; Length 88624;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 AATTGAAGTTATGTATCT 27  
|||||  
Db 49459 AATTGAAGTTATGTATCT 49477  
|||||

## RESULT 20

AC020676/8  
LOCUS 88624 bp DNA linear HTG 13-JUL-2000  
DEFINITION Homo sapiens clone RP11-8P23, LOW-PASS SEQUENCE SAMPLING.

AC020676  
AC020676.2 GI:9123792

HTG: HTGS\_PHASE0.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 88624)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-8P23

Unpublished

2 (bases 1 to 88624)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,

Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J.,

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,

Garbina,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lander,E., Lechoczky,J., Levine,R., Lien,C., Liu,G., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A. and Zody,M.

Direct Submission

Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:682672.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RN/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2995

Center clone name: 8\_P\_23

\* NOTE: This record contains 88 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1 882: contig of 882 bp in length  
\* 883 882: gap of 100 bp  
\* 886: contig of 904 bp in length  
\* 1887 1886: gap of 100 bp  
\* 1987 1905: contig of 919 bp in length  
\* 2306 3005: gap of 100 bp  
\* 3006 3003: contig of 898 bp in length  
\* 3304 4003: gap of 100 bp  
\* 4004 4923: contig of 922 bp in length  
\* 4926 5025: gap of 100 bp  
\* 5026 5924: contig of 899 bp in length  
\* 5925 6024: gap of 100 bp  
\* 6025 6937: contig of 913 bp in length  
\* 6938 7037: gap of 100 bp  
\* 7038 7949: contig of 912 bp in length  
\* 7950 8049: gap of 100 bp  
\* 8050 8920: contig of 871 bp in length  
\* 8921 9020: gap of 100 bp  
\* 9021 9894: contig of 874 bp in length  
\* 9895 9994: gap of 100 bp  
\* 9995 10908: contig of 914 bp in length  
\* 10909 11008: gap of 100 bp  
\* 11009 11921: contig of 913 bp in length  
\* 11922 12021: gap of 100 bp  
\* 12022 12931: contig of 910 bp in length  
\* 12932 13031: gap of 100 bp  
\* 13032 13947: contig of 916 bp in length  
\* 13948 14047: gap of 100 bp  
\* 14048 14949: contig of 902 bp in length  
\* 14950 15049: gap of 100 bp  
\* 15050 15960: contig of 911 bp in length  
\* 15961 16060: gap of 100 bp  
\* 16061 16971: contig of 911 bp in length  
\* 16972 17071: gap of 100 bp  
\* 17072 17996: contig of 925 bp in length  
\* 17997 18096: gap of 100 bp  
\* 18097 18986: contig of 890 bp in length  
\* 18987 19086: gap of 100 bp  
\* 19087 19992: contig of 906 bp in length  
\* 19993 20092: gap of 100 bp  
\* 20093 20958: contig of 866 bp in length  
\* 20959 21058: gap of 100 bp  
\* 21059 21948: contig of 890 bp in length  
\* 21949 22048: gap of 100 bp  
\* 22049 22957: contig of 909 bp in length  
\* 22958 23057: gap of 100 bp  
\* 23058 23967: contig of 910 bp in length  
\* 23968 24067: gap of 100 bp  
\* 24068 25001: contig of 934 bp in length  
\* 25002 25101: gap of 100 bp  
\* 25102 26000: contig of 899 bp in length  
\* 26001 26100: gap of 100 bp  
\* 26101 27013: contig of 913 bp in length  
\* 27014 27113: gap of 100 bp  
\* 27114 28031: contig of 918 bp in length  
\* 28032 28131: gap of 100 bp  
\* 28132 29041: contig of 910 bp in length  
\* 29042 29141: gap of 100 bp  
\* 29142 30038: contig of 897 bp in length  
\* 30039 30138: gap of 100 bp  
\* 30139 31031: contig of 893 bp in length  
\* 31032 31131: gap of 100 bp  
\* 31132 32049: contig of 918 bp in length  
\* 32050 32149: gap of 100 bp  
\* 32150 33064: contig of 915 bp in length  
\* 33065 33164: gap of 100 bp  
\* 33165 34070: contig of 906 bp in length  
\* 34071 34170: gap of 100 bp  
\* 34171 35094: contig of 924 bp in length  
\* 35095 35194: gap of 100 bp  
\* 35195 36126: contig of 932 bp in length  
\* 36127 36226: gap of 100 bp  
\* 36227 37155: contig of 929 bp in length

\* 37156 37255: gap of 100 bp  
\* 37256 38171: contig of 916 bp in length  
\* 38172 38271: gap of 100 bp  
\* 38272 39176: contig of 905 bp in length  
\* 39177 39276: gap of 100 bp  
\* 39277 40184: contig of 908 bp in length  
\* 40185 40284: gap of 100 bp  
\* 40285 41184: contig of 900 bp in length  
\* 41185 41284: gap of 100 bp  
\* 41285 42188: contig of 904 bp in length  
\* 42189 43177: contig of 889 bp in length  
\* 43178 43277: gap of 100 bp  
\* 43278 44199: contig of 922 bp in length  
\* 44200 44299: gap of 100 bp  
\* 44300 45219: contig of 920 bp in length  
\* 45220 45319: gap of 100 bp  
\* 45320 46209: contig of 890 bp in length  
\* 46210 46309: gap of 100 bp  
\* 46310 47208: contig of 899 bp in length  
\* 47209 47308: gap of 100 bp  
\* 47309 48233: contig of 925 bp in length  
\* 48334 49266: contig of 933 bp in length  
\* 49267 49366: gap of 100 bp  
\* 49367 50292: contig of 926 bp in length  
\* 50293 50392: gap of 100 bp  
\* 50393 51290: contig of 898 bp in length  
\* 51291 51390: gap of 100 bp  
\* 51391 52279: contig of 889 bp in length  
\* 52280 52379: gap of 100 bp  
\* 52380 53275: contig of 896 bp in length  
\* 53276 54289: contig of 914 bp in length  
\* 54290 54389: gap of 100 bp  
\* 54390 55311: contig of 922 bp in length  
\* 55312 55411: gap of 100 bp  
\* 55412 56319: contig of 908 bp in length  
\* 56320 56419: gap of 100 bp  
\* 56420 57325: contig of 906 bp in length  
\* 57326 57425: gap of 100 bp  
\* 57426 58342: contig of 917 bp in length  
\* 58343 58442: gap of 100 bp  
\* 58443 59358: contig of 916 bp in length  
\* 59359 59458: gap of 100 bp  
\* 59459 60380: contig of 922 bp in length  
\* 60381 61386: contig of 906 bp in length  
\* 60481 61486: gap of 100 bp  
\* 61387 62365: contig of 879 bp in length  
\* 62366 62465: gap of 100 bp  
\* 62466 63359: contig of 894 bp in length  
\* 63360 63459: gap of 100 bp  
\* 63460 64352: contig of 893 bp in length  
\* 64353 64452: gap of 100 bp  
\* 64453 65354: contig of 902 bp in length  
\* 65355 65454: gap of 100 bp  
\* 65455 66370: contig of 916 bp in length  
\* 66371 66470: gap of 100 bp  
\* 66471 67399: contig of 929 bp in length  
\* 67400 67499: gap of 100 bp  
\* 67500 68395: contig of 896 bp in length  
\* 68396 69391: contig of 896 bp in length  
\* 69392 69491: gap of 100 bp  
\* 69492 70404: contig of 913 bp in length  
\* 70405 70504: gap of 100 bp  
\* 70505 71416: contig of 912 bp in length  
\* 71417 71516: gap of 100 bp

Query Match 70.4%; Score 19; DB 2; Length 88624;  
Best Local Similarity 100.0%; Pred No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATCTATCCT 27  
 Db 78985 AATTGAAGTTATCTATCCT 78967

## RESULT 21

AF225899 106539 bp DNA linear PRI 23-FEB-2000  
 LOCUS Homo sapiens PAC clone 219d7, complete sequence.  
 DEFINITION AF225899  
 ACCESSION AF225899  
 VERSION AF225899.1 GI:7021528

## KEYWORDS

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 106539)  
 Fitzpatrick,E.S., Hammond,H.A., DeAngelis,D.M., Soderman,A.R.,  
 Wright,J.L., Liu,X., Larson,D., McGowan,J., Ziegler,S.,  
 Pritchard,L., Hess,J.F., Todd,J., Caskey,C.T. and Metzker,M.L.

## AUTHORS

Direct Submission  
 Submitted (18-JAN-2000) Department of Human Genetics, Merck & Co.,  
 Inc., Sumneytown Pike, West Point, PA 19486, USA

## FEATURES

Location/Qualifiers

1..106539

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="PAC 219d7"

complement (34..104)

/rpt\_family="MER5"

complement (384..662)

/rpt\_family="Alu"

1021..1304

/rpt\_family="Alu"

1800..2135

/rpt\_family="Alu"

2139..2439

/rpt\_family="Alu"

3214..3234

/rpt\_family="MLT2B2"

3588..3657

/rpt\_family="MLT2B2"

3982..4274

/rpt\_family="Alu"

5406..5550

/rpt\_family="MER3"

5444..5509

/rpt\_family="MER33"

complement (5558..5834)

/rpt\_family="Alu"

5852..5898

/rpt\_family="MER3"

complement (6534..6827)

/rpt\_family="Alu"

complement (8239..8568)

/rpt\_family="LTR12"

11864..12159

/rpt\_family="Alu"

complement (12838..13137)

/rpt\_family="Alu"

complement (14450..14691)

/rpt\_family="Alu"

complement (15004..15291)

/rpt\_family="Alu"

complement (16143..16422)

/rpt\_family="Alu"

complement (16813..17096)

/rpt\_family="Alu"

complement (17528..17706)

/rpt\_family="MIR"

21610..21890

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

/rpt\_family="Alu"  
 23049..23329  
 /rpt\_family="Alu"  
 24379..24585  
 /rpt\_family="L1"  
 complement (27993..28279)  
 /rpt\_family="Alu"  
 28842..28933  
 /rpt\_family="MIR"  
 30439..30783  
 /note="L18270 Human chromosome 2 STS UT426"  
 complement (31317..31610)  
 /rpt\_family="Alu"  
 32555..37857  
 /rpt\_family="L1"  
 33985..34281  
 /rpt\_family="Alu"  
 35580..35863  
 /rpt\_family="Alu"  
 complement (37874..38708)  
 /rpt\_family="L1"  
 38616..38712  
 /rpt\_family="MIR"  
 complement (41075..41331)  
 /rpt\_family="Alu"  
 complement (42064..42184)  
 /rpt\_family="Alu"  
 42309..42413  
 /rpt\_family="MIR"  
 complement (42689..42777)  
 /rpt\_family="MIR"  
 43887..44156  
 /rpt\_family="Alu"  
 46963..47240  
 /rpt\_family="Alu"  
 complement (47837..48028)  
 /rpt\_family="MIR"  
 complement (51874..52142)  
 /rpt\_family="Alu"  
 52754..52801  
 /rpt\_family="MER2"  
 58298..58529  
 /rpt\_family="Alu"  
 58600..58721  
 /rpt\_family="L1"  
 complement (59586..59701)  
 /rpt\_family="Alu"  
 60048..60335  
 /rpt\_family="Alu"  
 60964..61242  
 /rpt\_family="Alu"  
 complement (62514..63136)  
 /rpt\_family="Alu"  
 complement (68228..68344)  
 /rpt\_family="MER20"  
 69944..70696  
 /note="G06727 human STS WI-7843"  
 70776..71041  
 /rpt\_family="Alu"  
 75420..75585  
 /rpt\_family="Alu"  
 complement (77681..77951)  
 /rpt\_family="Alu"  
 complement (78093..78393)  
 /rpt\_family="Alu"  
 79977..80042  
 /rpt\_family="THE1"  
 complement (81069..81363)  
 /rpt\_family="Alu"  
 81561..81913  
 /note="G09915 human STS CHLC.GGAA 19E07.PI7317 clone  
 GGAA19E07"  
 complement (82076..82380)  
 repeat\_region

```

repeat_region      /rpt_family="Alu"
82595. .82857
/rpt_family="Alu"
complement(84782. .85059)
/rpt_family="Alu"
complement(86131. .86655)
/rpt_family="Alu"
complement(86698. .86772)
/rpt_family="MLT1"
87328. .87517
/rpt_family="Alu"
complement(88025. .88300)
/rpt_family="Alu"
complement(88382. .88537)
/rpt_family="Alu"
complement(89103. .92131)
/rpt_family="L1"
complement(89607. .89870)
/rpt_family="Alu"
complement(92183. .92468)
/rpt_family="Alu"
complement(92638. .92723)
/rpt_family="L1"
complement(93254. .93542)
/rpt_family="Alu"
93950. .94195
/rpt_family="MER43"
94819. .94882
/rpt_family="MER43"
97105. .97472
/rpt_family="THE1"
99270. .99669
/notes="G08815 human STS CHLC.GGAA19B07.P17317 CLONE
GGAA19B07"
repeat_region      complement(101199. .101438)
/rpt_family="Alu"
101753. .102038
/rpt_family="Alu"
complement(102216. .102638)
/rpt_family="LOR1"
complement(103792. .104101)
/rpt_family="L1"
complement(104170. .104465)
/rpt_family="Alu"
104528. .104642
/rpt_family="L1"
105250. .105545
/rpt_family="Alu"
105891. .106161
/rpt_family="Alu"
ORIGIN
Query Match      70.4%; Score 19; DB 9; Length 106539;
Best Local Similarity 100.0%; Pred.No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 AATTGAAGTTATGTCCT 27
|||||
Db 61946 AATTGAAGTTATGTCCT 61964
|||||

RESULT 22
AC103881
LOCUS      AC103881      108100 bp      DNA      linear      PRI 23-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-638D14 from 2, complete sequence.
ACCESSION  AC103881
VERSION     AC103881.4 GI:19697537
KEYWORDS   HTG.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 108100)

```

AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
9847074  
2 (bases 1 to 108100)  
Levy, A., Kozlowski, A. and Spalding, L.  
The sequence of Homo sapiens BAC clone RP11-638D14  
Unpublished (2001)  
3 (bases 1 to 108100)  
Waterston, R.H.  
Direct Submission  
Submitted (29-NOV-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 108100)  
Waterston, R.  
Direct Submission  
Submitted (23-MAR-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 23, 2002 this sequence version replaced gi:18030160.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0638D14  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-789A18, 2000 bp overlap;  
the clone sequenced to the right is RP11-515K12, 2000 bp overlap.  
Actual start of this clone is at base position 146042 of  
RP11-789A18; actual end is at base position 23686 of RP11-515K12.  
Polymorphisms have been identified between AC093113 and AC103881.

A transposon was identified in a growth of the clone RP11-638D14,  
which is not a part of the submitted sequence.

FEATURES  
source  
Location/Qualifiers  
1. .108100  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

```
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-638D14"
/clone_lib="RPC1-11"
411..867
/rpt_family="MaLR"
repeat_region
1022..1329
/rpt_family="Alu"
repeat_region
3858..4066
/rpt_family="MIR"
repeat_region
4074..4179
/rpt_family="MIR"
repeat_region
4527..4695
/rpt_family="MIR"
repeat_region
4774..7112
/rpt_family="L1"
repeat_region
7123..7652
/rpt_family="L1"
repeat_region
7629..7771
/rpt_family="L1"
repeat_region
8144..8445
/rpt_family="Alu"
repeat_region
8589..8595
/rpt_family="AT_rich"
repeat_region
10489..10564
/rpt_family="MaLR"
repeat_region
10566..10845
/rpt_family="MaLR"
repeat_region
10846..11208
/rpt_family="MaLR"
repeat_region
11212..11985
/rpt_family="L1"
repeat_region
11987..12299
/rpt_family="Alu"
repeat_region
12300..12779
/rpt_family="L1"
repeat_region
12789..13247
/rpt_family="L1"
repeat_region
13277..13669
/rpt_family="MaLR"
repeat_region
13705..13828
/rpt_family="L1"
repeat_region
13829..14096
/rpt_family="MER2_type"
repeat_region
14095..14190
/rpt_family="MER2_type"
repeat_region
14191..14346
/rpt_family="L1"
repeat_region
14783..14868
/rpt_family="AT_rich"
repeat_region
15037..15357
/rpt_family="Alu"
repeat_region
16001..16141
/rpt_family="L2"
repeat_region
16207..16245
/rpt_family="MIR"
repeat_region
16734..16773
/rpt_family="(CAAAA)n"
repeat_region
16922..17761
/rpt_family="L1"
repeat_region
17933..17957
/rpt_family="(CA)n"
repeat_region
17980..18101
/rpt_family="MIR"
repeat_region
18102..18465
/rpt_family="MaLR"
repeat_region
18466..18516
/rpt_family="MIR"
repeat_region
19240..19313
/rpt_family="CT-rich"
repeat_region
21201..21295
/rpt_family="L2"
```

```
repeat_region 21297..21835
/rpt_family="ERVL"
repeat_region 27448..27997
/rpt_family="ERVL"
repeat_region 28054..28091
/rpt_family="AT_rich"
repeat_region 28096..28202
/rpt_family="Alu"
repeat_region 28210..28246
/rpt_family="(TAAA)n"
repeat_region 28296..28319
/rpt_family="(TG)n"
repeat_region 28324..28779
/rpt_family="L1"
repeat_region 28780..29214
/rpt_family="MaLR"
repeat_region 29215..29704
/rpt_family="L1"
repeat_region 29707..30227
/rpt_family="L1"
repeat_region 30216..30997
/rpt_family="L1"
repeat_region 31381..31798
/rpt_family="L1"
repeat_region 31813..31965
/rpt_family="MER1_type"
repeat_region 33297..33558
/rpt_family="CR1"
repeat_region 34871..34894
/rpt_family="AT_rich"
repeat_region 41290..41408
/rpt_family="MIR"
repeat_region 41745..42317
/rpt_family="ERVL"
repeat_region 42548..42714
/rpt_family="MIR"
repeat_region 42715..42737
/rpt_family="AT_rich"
repeat_region 44278..44461
/rpt_family="MIR"
repeat_region 44508..44542
/rpt_family="AT_rich"
repeat_region 45741..45799
/rpt_family="MaLR"

Query Match 70.4%; Score 19; DB 9; Length 108100;
Best Local Similarity 81.5%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGATCCT 27
Db 29004 GTGGCAGAGATTGAAGTTATGATCAT 29030

RESULT 23
CR407559
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
CR407559 148587 bp DNA linear HTG 11-AUG-2004
Danio rerio clone CH211-283G2, WORKING DRAFT SEQUENCE.
CR407559
CR407559.7 GI:50978507
HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 148587)
Lovell, J.
Direct Submission
Submitted (10-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT
On Aug 4, 2004 this sequence version replaced gi:50724958.
```



```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZC283C2
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 148583 bases at least Q40
Consensus quality: 148587 bases at least Q30
Consensus quality: 148587 bases at least Q20
Insert size: 148587; sum-of-contigs
Insert size: 163547; 10.1% error; agarose-fp
Quality coverage: 11.56x in Q20 bases; sum-of-contigs Quality
coverage: 11.33x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
-----
Location/Qualifiers
1..148587
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-283G2"
/clone_lib="CHORI-211"
misc_feature 1..148587
/note="assembly_fragment:01416"

ORIGIN
Query Match 70.4%; Score 19; DB 2; Length 148587;
Best Local Similarity 81.5%; Pred No. 3.2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATCTCT 27
||||| ||||| ||||| |||||
Db 54365 GCGCGCGCAATTGAATCTATGCAACCT 54391

RESULT 24
HSXKSRPXM
LOCUS
DEFINITION
Homo sapiens chromosome X sequence from 5 cosmids, map Xp21.1
region DXS709-SRPX, complete sequence.
ACCESSION
AL121577
VERSION
AL121577.1 GI:5931934
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ramsey J., Borzym, K., Steffens, C., Klein, M., Monaco, A., Ho, M.,
Reinhardt, R. and Lehrach, H.
Unpublished
2 (bases 1 to 175531)
MPIMG.
Direct Submission
Submitted (27-SEP-1999) MPIMG, Abt. Lehrach, Max Planck Institut
Fuer Molekulare Genetik, Innestrasse 73, Berlin, 14195 Germany
contig 01 1..175531
Clone ICRF104 3C29 received from the Resource Centre of the Human
Genome Project at the Max-Planck-Institut for Molecular Genetics.
Clones ICRF104 135D2, ICRF104 203D3, ICRF104 47E10, and ICRF104
49A7 received from A. Monaco.
Location/Qualifiers
1..175531
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
/chromosome="X"
/map="Xp21.1"
/clone="cosmid: ICRF104 3C29"
/clone_lib="ICRF-Library: L4/FSC X, originating institute:
Imperial Cancer Research Fund (ICRF), creator: Dean
Nizetic"
/note="region between markers DXS709-SRPX"
1..44226
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="3C29"
35604..72266
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="135D2"
61035..103556
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="203D3"
94955..134676
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="47E10"
128388..175531
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="49A7"
complement(1297..1348)
/note="MLTIG"
/rpt_family="LTR/MaLR"
2391..2786
/note="LINE2"
/rpt_family="LINE/L2"
2824..3370
/note="MER68A"
/rpt_family="Other/MER21_gro"
3379..3484
/note="LINE2"
/rpt_family="LINE/L2"
complement(5151..5336)
/note="MERSA"
/rpt_family="DNA/MER1_type"
complement(6836..6964)
/note="MLTIG"
/rpt_family="LTR/MaLR"
8576..8766
/note="MIR"
/rpt_family="SINE/MIR"
complement(8784..8810)
/note="(TAAAA)n"
/rpt_family="Simple repeat"
complement(8813..9120)
/note="AluSg"
/rpt_family="SINE/Alu"
9150..9258
/note="MERSA"
/rpt_family="DNA/MER1_type"
complement(9551..9663)
/note="LINE2"
/rpt_family="LINE/L2"
9949..10244
/note="LINE2"
/rpt_family="LINE/L2"
complement(10276..10679)
/note="L1ME2"
/rpt_family="LINE/L1"
11097..11273
/note="MER5B"

```

```
repeat_region /rpt_family="DNA/MER1_type"
complement(11819..12052)
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
complement(12106..12208)
/note="MER5A"
repeat_region /rpt_family="DNA/MER1_type"
complement(12211..12377)
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
12780..12808
/note="AT rich"
repeat_region /rpt_family="Low_complexity"
13211..13231
/note="AT rich"
repeat_region /rpt_family="Low_complexity"
13675..13842
/note="MER5A"
repeat_region /rpt_family="DNA/MER1_type"
complement(15296..15743)
/note="MLT1C"
repeat_region /rpt_family="LTR/MaLR"
complement(20444..20557)
/note="(TAAA)n"
repeat_region /rpt_family="Simple repeat"
complement(20723..20977)
/note="L1MD2"
repeat_region /rpt_family="LINE/L1"
complement(21024..21308)
/note="L1P5"
repeat_region /rpt_family="LINE/L1"
21434..21464
/note="AT rich"
repeat_region /rpt_family="Low_complexity"
22607..22671
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
23030..23106
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
24353..24442
/note="MER5A"
repeat_region /rpt_family="DNA/MER1_type"
25051..25308
/note="L1MC4"
repeat_region /rpt_family="LINE/L1"
complement(25337..31471)
/note="L1PA2"
repeat_region /rpt_family="LINE/L1"
31474..31823
/note="L1MC4"
repeat_region /rpt_family="LINE/L1"
complement(32064..32118)
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
32363..32389
/note="AT rich"
repeat_region /rpt_family="Low_complexity"
complement(32409..32515)
/note="L1MA5A"
repeat_region /rpt_family="LINE/L1"
32514..32678
/note="L1MA5A"
repeat_region /rpt_family="LINE/L1"
complement(32689..33354)
/note="L1PA15"
repeat_region /rpt_family="LINE/L1"
34927..35033
/note="(GA)n"
repeat_region /rpt_family="Simple repeat"
complement(35511..36012)
/note="L1ME3"
/rpt_family="LINE/L1"

repeat_region complement(36042..36210)
/note="(CATA)n"
/rpt_family="Simple repeat"
complement(36700..36731)
/note="(CAAAA)n"
repeat_region /rpt_family="Simple repeat"
complement(37034..37386)
/note="THE1C"
/rpt_family="LTR/MaLR"
37387..37472
/note="LINE2"
repeat_region /rpt_family="LINE/L2"
38142..38176
/note="AT rich"
repeat_region /rpt_family="Low_complexity"
complement(38203..38260)
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
complement(40343..40406)
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
complement(40728..40966)
/note="MIR"
repeat_region /rpt_family="SINE/MIR"
41055..41237

Query Match 70.4%; Score 19; DB 9; Length 175531;
Best Local Similarity 81.5%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATCCT 27
DB 140719 GCTTCTCATTTGAAGTTTGTATCCT 140745

RESULT 25
AC125238 196622 bp DNA linear PRI 15-OCT-2002
LOCUS Homo sapiens BAC clone RP11-711C24 from 2, complete sequence.
DEFINITION AC125238
ACCESSION AC125238
VERSION AC125238.5 GI:22267884
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 196622)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 196622)
AUTHORS Nguyen,C., Bielicki,L. and Schatzkamer,K.
TITLE The sequence of Homo sapiens BAC clone RP11-711C24
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 196622)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 196622)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 196622)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

6 (bases 1 to 196622)  
Waterston,R.  
Direct Submission  
Submitted (15-OCT-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Aug 16, 2002 this sequence version replaced gi:22138709.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0711C24  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male  
donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,  
Tateno,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is  
overlapped by AC069314 and AC010138.

A transposon was identified in the finished region of this clone  
and removed prior to submission. The transposon would insert after  
base pair 83916 of this sequence.

The sequence from base 156374 to base 156453 was derived from one  
plasmid subclone.

The sequence from base 156479 to base 156499 was derived from one  
plasmid subclone.

Polymorphisms have been identified between AC069314, AC010138 and  
this sequence.

Data from AC069314 and AC010138 was used to finish this clone.

FEATURES  
source

Location/Qualifiers  
1. .196622  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-711C24"  
/clone\_lib="RPCT-11"  
repeat\_region 215..236  
/rpt\_family="(TTTAA)n"

repeat\_region 240..520  
/rpt\_family="Alu"  
repeat\_region 892..1086  
/rpt\_family="L1"  
repeat\_region 1575..1885  
/rpt\_family="Alu"  
repeat\_region 1991..2078  
/rpt\_family="MIR"  
repeat\_region 2079..2387  
/rpt\_family="Alu"  
repeat\_region 2418..2487  
/rpt\_family="Alu"  
repeat\_region 2488..2764  
/rpt\_family="Alu"  
repeat\_region 2771..2871  
/rpt\_family="MIR"  
repeat\_region 3123..3190  
/rpt\_family="L2"  
repeat\_region 3296..3599  
/rpt\_family="Alu"  
repeat\_region 3612..3919  
/rpt\_family="Alu"  
repeat\_region 3926..3988  
/rpt\_family="MIR"  
repeat\_region 4192..4493  
/rpt\_family="Alu"  
repeat\_region 5165..5256  
/rpt\_family="ACHobo"  
repeat\_region 5257..5593  
/rpt\_family="L1"  
repeat\_region 5599..5623  
/rpt\_family="(TG)n"  
repeat\_region 5624..5748  
/rpt\_family="Alu"  
repeat\_region 5752..5929  
/rpt\_family="L1"  
repeat\_region 6043..6552  
/rpt\_family="ACHobo"  
repeat\_region 6670..6803  
/rpt\_family="L2"  
repeat\_region 6804..7097  
/rpt\_family="Alu"  
repeat\_region 7098..7192  
/rpt\_family="L2"  
repeat\_region 7193..7515  
/rpt\_family="Alu"  
repeat\_region 7516..7745  
/rpt\_family="L2"  
repeat\_region 7746..8054  
/rpt\_family="Alu"  
repeat\_region 8055..8264  
/rpt\_family="L2"  
repeat\_region 8265..8565  
/rpt\_family="Alu"  
repeat\_region 8566..9242  
/rpt\_family="L2"  
repeat\_region 9381..9682  
/rpt\_family="Alu"  
repeat\_region 9688..9997  
/rpt\_family="Alu"  
repeat\_region 10131..10384  
/rpt\_family="L2"  
repeat\_region 10773..11075  
/rpt\_family="Alu"  
repeat\_region 11076..11379  
/rpt\_family="Alu"  
repeat\_region 11492..11580  
/rpt\_family="MIR"  
repeat\_region 11927..12020  
/rpt\_family="MIR"  
repeat\_region 12146..12237  
/rpt\_family="MER2\_type"  
repeat\_region 12238..12534

```

repeat_region      /rpt family="Alu"
12535 12682
/rpt family="MER2_type"
12683 13170
/rpt family="MaLR"
13171 13193
/rpt family="MER2_type"
13621 14034
/rpt family="MaLR"
14683 14966
/rpt family="L1"
14967 15250
/rpt family="Alu"
15251 15303
/rpt family="L1"
15796 16074
/rpt family="Alu"
16075 16315

Query Match      70.4% Score 19; DB 9; Length 196622;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
|||||
Db 119311 AATTGAAGTTATGATCCT 119329

RESULT 26
AC094790
LOCUS      AC094790 235811 bp DNA linear HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-4C5, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION  AC094790 GI:30466636
VERSION     AC094790.7
KEYWORDS    HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 235811)
Muzny,D,Marie, Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Altschrocks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,I.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,I., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseged,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahandartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,K., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

```

# TITLE JOURNAL REFERENCE

Unpublished  
2 (bases 1 to 235811)

Worley,K.C.

Direct Submission

## TITLE JOURNAL

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 235811)

Rat Genome Sequencing Consortium.

Direct Submission

## TITLE JOURNAL

Submitted (09-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On May 9, 2003 this sequence version replaced gi:24818376.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GBUU

Center clone name: CH230-4C5

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 227691 bases at least Q40

Consensus quality: 228978 bases at least Q30

Consensus quality: 229886 bases at least Q20

Estimated insert size: 241168; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 234674: contig of 234674 bp in length

\* 234675 234774: gap of unknown length

[illegible]

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tatenio in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

#### FEATURES

source	Location/Qualifiers
1..200339	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="15"
	/map="15"
	/clone="RP23-181J12"
	/clone_lib="RPCI-23"
1..1573	/rpt_family="L1"
1572..2347	/rpt_family="L1"
2341..2875	/rpt_family="L1"
2959..3167	/rpt_family="L1"
3432..3580	/rpt_family="L1"
3632..3765	/rpt_family="MaLR"
3771..4057	/rpt_family="L1"
4209..4277	/rpt_family="B2"
4565..4691	/rpt_family="MaLR"
5231..5327	/rpt_family="MaLR"
5394..5489	/rpt_family="ERV1"
6468..6527	/rpt_family="ERV1"
6839..6987	/rpt_family="Alu"
8635..9669	/rpt_family="L1"
9669..9898	/rpt_family="L1"
10122..10208	/rpt_family="MaLR"
10291..10748	/rpt_family="RMER15"
10911..10979	/rpt_family="MaLR"
11369..11600	

repeat_region	/rpt_family="MaLR"
repeat_region	/rpt_family="RMER13A"
repeat_region	/rpt_family="Alu"
repeat_region	12714..13125
repeat_region	/rpt_family="L1"
repeat_region	13152..13269
repeat_region	/rpt_family="L1"
repeat_region	13786..13863
repeat_region	/rpt_family="L2"
repeat_region	13864..14082
repeat_region	/rpt_family="MER1_type"
repeat_region	14228..14307
repeat_region	/rpt_family="MaLR"
repeat_region	14385..14519
repeat_region	/rpt_family="Alu"
repeat_region	14525..14802
repeat_region	/rpt_family="MaLR"
repeat_region	15129..15431
repeat_region	/rpt_family="MaLR"
repeat_region	15795..15941
repeat_region	/rpt_family="L1"
repeat_region	16559..16817
repeat_region	/rpt_family="L1"
repeat_region	17585..17932
repeat_region	/rpt_family="ERVK"
repeat_region	18027..18300
repeat_region	/rpt_family="B4"
repeat_region	18399..18732
repeat_region	/rpt_family="ERVK"
repeat_region	21585..22438
repeat_region	/rpt_family="ERVK"
repeat_region	22441..22774
repeat_region	/rpt_family="ERVK"
repeat_region	24311..24535
repeat_region	/rpt_family="MaLR"
repeat_region	24906..25152
repeat_region	/rpt_family="B4"
repeat_region	25951..26021
repeat_region	/rpt_family="B4"
repeat_region	27923..28031
repeat_region	/rpt_family="Alu"
repeat_region	28635..28756
repeat_region	/rpt_family="RMER15"
repeat_region	28774..28921
repeat_region	/rpt_family="B2"
repeat_region	28940..29227
repeat_region	/rpt_family="L1"
repeat_region	29257..29388
repeat_region	/rpt_family="L1"
repeat_region	34736..34905
repeat_region	/rpt_family="B2"
repeat_region	35031..35378
repeat_region	/rpt_family="L1"
repeat_region	36446..36534
repeat_region	/rpt_family="Alu"
repeat_region	36535..36918
repeat_region	/rpt_family="ERVK"
repeat_region	36919..36980
repeat_region	/rpt_family="Alu"
repeat_region	37306..37382
repeat_region	/rpt_family="ERVK"
repeat_region	37682..38305
repeat_region	/rpt_family="ERVK"
repeat_region	38438..38670
repeat_region	/rpt_family="ERVK"
repeat_region	38916..39736
repeat_region	/rpt_family="ERVK"
repeat_region	39750..40236
repeat_region	/rpt_family="L1"
repeat_region	40249..40443
repeat_region	/rpt_family="ERVK"

```

repeat_region 40619..40819
/rpt_family="Maur"
repeat_region 41238..41549
/rpt_family="L1"
repeat_region 41861..42144
/rpt_family="RMER3"
repeat_region 42610..43314
/rpt_family="ERVK"
repeat_region 43354..43450
/rpt_family="Alu"

Query Match 68.9%; Score 18.6; DB 10; Length 200339;
Best Local Similarity 84.0%; Pred. No. 5.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GCGCCGAATGAGTTATGATCTCT 27
|||||
Db 113728 GCGCCAGTAAGAATTATGATGCT 113704

RESULT 29
AC121459/c
LOCUS
DEFINITION AC121459 218328 bp DNA linear HTG 09-OCT-2002
Rattus norvegicus clone CH230-245A4, *** SEQUENCING IN PROGRESS
*** 2 unordered pieces.
ACCESSION AC121459
VERSION AC121459.4 GI:22856081
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 218328)
Muzny,D,Marie, Metzker,M,Lee, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Gantä,R., Garcia,A., Garner,T., Garza,M.,
Gebregregis,E., Geet,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kowar,C.,
Kows,C., Kratt,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Loreshuwa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Flopper,F., Polndexter,A., Popovic,D., Primus,E., Fu,L.-L.,
Fuazo,M., Quintero,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

```

```

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 218328)
Worley,K.C.
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 218328)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 14, 2002 this sequence version replaced gi:22450321.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYKW
Center clone name: CH230-245A4
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 171033 bases at least Q40
Consensus quality: 174532 bases at least Q30
Consensus quality: 176624 bases at least Q20
Estimated insert size: 189733; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 217012: contig of 217012 bp in length
* 217013 217112: gap of unknown length
* 217113 218328: contig of 1216 bp in length.
Location/Qualifiers
1. 218328
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-245A4"
1. 2059
/note="wgs_contig"
misc_feature
19850..20857
/note="wgs_contig"
misc_feature
21359..22735
/note="wgs_contig"
FEATURES
source
misc_feature
misc_feature
misc_feature

```







Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 227929)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On May 9, 2003 this sequence version replaced gi:22771587.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRBY  
 Center clone name: CH230-4N15  
 ----- Summary Statistics  
 Assembly program: Atlas  
 Consensus quality: 200503 bases at least Q40  
 Consensus quality: 205667 bases at least Q30  
 Consensus quality: 209426 bases at least Q20  
 Estimated insert size: 217540; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 ----- NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as runs of N, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
 ----- This sequence will be replaced  
 ----- by the finished sequence as soon as it is available and  
 ----- the accession number will be preserved.  
 \* 1 227929: contig of 227929 bp in length.

FEATURES  
 source  
 Location/Qualifiers  
 1..227929  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-4N15"  
 1..3298  
 /note="wgs end extension  
 clone\_end:Sp6"  
 5495..6182  
 /note="clone boundary  
 clone\_end:Sp6  
 site:Scor1  
 end\_sequence:BH309552"

misc\_feature  
 1..3298  
 /note="wgs end extension  
 clone\_end:Sp6"  
 5495..6182  
 /note="clone boundary  
 clone\_end:Sp6  
 site:Scor1  
 end\_sequence:BH309552"

misc\_feature  
 1..3298  
 /note="wgs end extension  
 clone\_end:Sp6"  
 5495..6182  
 /note="clone boundary  
 clone\_end:Sp6  
 site:Scor1  
 end\_sequence:BH309552"

ORIGIN  
 Query Match 68.9%; Score 18.6; DB 2; Length 227929;  
 Best Local Similarity 84.0%; Pred. No. 5.4e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 3 GGCCGAATTGAAGTTATGATCT 27  
 131881 GGCTACAATTGAAGTTTGTCT 131887

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

RESULT 32  
 AC109678  
 LOCUS  
 DEFINITION  
 Rattus norvegicus clone CH230-89E12, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 8 unordered pieces.  
 AC109678.5 GI:24818582  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 236612)  
 Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Chen,Y., Chen,Z.,  
 Chacko,J., Chavez,D., Chen,R., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Kocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hernandez,J.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jarkatshy,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowas,C., Kraft,C.I., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorenshew,L., Loulseg,H., Lozano,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,B., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwakolameh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,J., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,E., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 236612)  
 Worley,K.C.  
 Direct Submission  
 Submitted (07-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 236612)  
 Rat Genome Sequencing Consortium.



Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Scojanovic, N., Stubbs, M., Talanas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE JOURNAL

Submitted (11-AUG-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 11, 2004 this sequence version replaced gi:51036443.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@broad.mit.edu

----- Project Information

Center project name: L14656

Center clone name: 74\_L\_20

\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 16605: contig of 16605 bp in length  
\* 16606: gap of unknown length  
\* 16706: contig of 19645 bp in length  
\* 36350: contig of 19645 bp in length  
\* 36450: gap of unknown length  
\* 36351: gap of unknown length  
\* 36451: contig of 23588 bp in length  
\* 60039: gap of unknown length  
\* 60138: contig of 5791 bp in length  
\* 65929: gap of unknown length  
\* 66030: contig of 29072 bp in length  
\* 95101: gap of unknown length  
\* 95201: contig of 32229 bp in length  
\* 95202: gap of unknown length  
\* 127431: contig of 5947 bp in length  
\* 127530: gap of unknown length  
\* 133477: contig of 5947 bp in length  
\* 133478: gap of unknown length  
\* 168446: contig of 34869 bp in length  
\* 168447: gap of unknown length  
\* 168547: contig of 78554 bp in length.

## FEATURES source

Location/Qualifiers  
1..247100  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/map="5"  
/clones="RP23-74L20"  
/clone\_lib="RPCI-23 Female Mouse BAC"

## ORIGIN

Query Match 68.9%; Score 18.6; DB 2; Length 247100;  
Best Local Similarity 84.0%; Pred. No. 5.4e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 3 GGCGGCAATTGAAGTTATGTATCT 27

Db 161941 GGACGATTGAGTTATGTATCTCT 161965

## RESULT 34 AC129421/c LOCUS

DEFINITION Rattus norvegicus clone CH230-62C13, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 5 unordered pieces.

## ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AC129421  
AC129421.3 GI:515138103  
HTG: HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 290066)

Muzny, D. Marie., Zetser, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalto, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, M., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, M., Hollins, B., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Huijck, S., Huijck, S., Hume, J., Idlerbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mancum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokeme, O., Okunolu, G., Olarnpungsoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, F., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 290066)

Worley, K.C.

Direct Submission

Submitted (30-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 290066)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23267370.

## COMMENT



of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 5 (bases 1 to 146945)  
 REFERENCES  
 AUTHORS  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 6 (bases 1 to 146945)  
 REFERENCES  
 AUTHORS  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (31-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jan 31, 2003 this sequence version replaced gi:6492471.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht>

#### FEATURES

source  
 1..146945  
 /location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /clone="RP11-19B18"  
 757..883  
 /standard\_name="WI-16568"  
 851..966  
 /standard\_name="A010A40"  
 2810..2839  
 /rpt\_family="(A)n"  
 3475..3512  
 /rpt\_family="T-rich"  
 3507..3632  
 /standard\_name="WI-13379"  
 3511..3648  
 /standard\_name="WI-14321"  
 complement(4059..4362)  
 /rpt\_family="AluSc"  
 repeat\_region  
 repeat\_region  
 STS  
 STS  
 repeat\_region

repeat\_region  
 4890..4922  
 /rpt\_family="AT-rich"  
 repeat\_region  
 5231..5853  
 /rpt\_family="LIME3A"  
 repeat\_region  
 5963..6237  
 /rpt\_family="AluSg"  
 repeat\_region  
 6797..7094  
 /rpt\_family="AluJo"  
 repeat\_region  
 7304..7614  
 /rpt\_family="AluY"  
 repeat\_region  
 7630..8065  
 /rpt\_family="HAL1"  
 repeat\_region  
 8132..8366  
 /rpt\_family="HAL1"  
 repeat\_region  
 8413..8724  
 /rpt\_family="AluSx"  
 repeat\_region  
 8868..8941  
 /rpt\_family="HAL1"  
 repeat\_region  
 9016..9163  
 /rpt\_family="(TA)n"  
 complement(9164..9450)  
 /rpt\_family="AluY"  
 repeat\_region  
 complement(9476..9781)  
 /rpt\_family="AluSc"  
 repeat\_region  
 9785..9805  
 /rpt\_family="AT-rich"  
 complement(9813..9848)  
 /rpt\_family="LIMB8"  
 repeat\_region  
 complement(9849..10139)  
 /rpt\_family="AluSx"  
 repeat\_region  
 complement(10140..10277)  
 /rpt\_family="LIMB8"  
 repeat\_region  
 complement(10304..10680)  
 /rpt\_family="LIMB3"  
 repeat\_region  
 10681..10927  
 /rpt\_family="AluJ"  
 repeat\_region  
 10933..10984  
 /rpt\_family="LIME1"  
 complement(10990..11244)  
 /rpt\_family="LIMB3"  
 repeat\_region  
 complement(11249..11871)  
 /rpt\_family="L1"  
 repeat\_region  
 complement(12081..12419)  
 /rpt\_family="THE1B"  
 repeat\_region  
 complement(12687..12937)  
 /rpt\_family="L1"  
 repeat\_region  
 12941..13226  
 /rpt\_family="AluSg"  
 repeat\_region  
 complement(13275..13393)  
 /rpt\_family="L2"  
 repeat\_region  
 complement(13526..13990)  
 /rpt\_family="L2"  
 repeat\_region  
 14980..15014  
 /rpt\_family="AT-rich"  
 repeat\_region  
 15365..15385  
 /rpt\_family="(GA)n"

Query Match 68.1%; Score 18.4; DB 9; Length 146945;  
 Best Local Similarity 95.0%; Pred. No. 6.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CAATTGAAGTTATGTTATCCT 27  
 |||||  
 Db 43751 CAATTGAGGTTATGTTATCCT 43732

#### RESULT 36

AC110187/c AC110187 180254 bp DNA linear HTG 13-MAR-2004  
 LOCUS Mus musculus chromosome 7 clone RP24-541E23 map 7, WORKING DRAFT  
 DEFINITION SEQUENCE, 18 unordered pieces.  
 ACCESSION AC110187  
 VERSION AC110187.5 GI:45430174

KEYWORDS  
SOURCE

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Mus musculus (house mouse)

## MUSCULUS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 180254)

## AUTHORS

Birren,B., Nusbaum,C. and Lander,E.

## JOURNAL

Mus musculus chromosome 7, clone RP24-541E23

## REFERENCE

Unpublished  
2 (bases 1 to 180254)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J.J., Chazaro,B.,  
Chospel,I., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,  
Landers,T., Lehocsky,J., Levine,R., Liu,G., Maclean,C.,  
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,  
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosatti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (10-FEB-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

## AUTHORS

3 (bases 1 to 180254)  
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J.J., Choepel,Y.,  
Collimore,A., Cook,A., Cooke,P., Corum,B., DeArelano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagoopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

Submitted (13-MAR-2004) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Mar 13, 2004 this sequence version replaced gi:20503122.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L19479  
Center Clone name: 541\_E\_23

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 174704 bases at least Q40  
Consensus quality: 177216 bases at least Q30  
Consensus quality: 178020 bases at least Q20  
Insert size: 178000; agarose-fp  
Insert size: 178494; sum-of-contigs  
Quality coverage: 8.0 in Q20 bases; agarose-fp  
Quality coverage: 8.0 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1910: contig of 1910 bp in length  
\* 1911 2010: gap of 100 bp  
\* 2011 25284: contig of 23274 bp in length  
\* 25285 25384: gap of 100 bp  
\* 25385 30044: contig of 4660 bp in length  
\* 30045 30144: gap of 100 bp  
\* 30145 38883: contig of 8739 bp in length  
\* 38884 38983: gap of 100 bp  
\* 38984 62614: contig of 23630 bp in length  
\* 62615 62713: gap of 100 bp  
\* 62714 64667: contig of 1954 bp in length  
\* 64668 64767: gap of 100 bp  
\* 64768 66216: contig of 1449 bp in length  
\* 66217 66316: gap of 100 bp  
\* 66317 75523: contig of 9207 bp in length  
\* 75524 75623: gap of 100 bp  
\* 75624 78993: contig of 3370 bp in length  
\* 78994 79093: gap of 100 bp  
\* 79094 95809: contig of 16716 bp in length  
\* 95810 95909: gap of 100 bp  
\* 95910 98133: contig of 2224 bp in length  
\* 98134 98233: gap of 100 bp  
\* 98234 99926: contig of 1693 bp in length  
\* 99927 100026: gap of 100 bp  
\* 100027 122553: contig of 22527 bp in length  
\* 122554 122653: gap of 100 bp  
\* 122654 142576: contig of 19923 bp in length  
\* 142577 142676: gap of 100 bp  
\* 142677 157739: contig of 15063 bp in length  
\* 157740 157839: gap of 100 bp  
\* 157840 167044: contig of 9205 bp in length  
\* 167045 167144: gap of 100 bp  
\* 167145 170812: contig of 3668 bp in length  
\* 170813 170912: gap of 100 bp  
\* 170913 180254: contig of 9342 bp in length.

## FEATURES

## source

1..180254  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="7"  
/map="7"  
/clone="RP24-541E23"  
/clone\_lib="RPCI-24 Male Mouse BAC"

## misc\_feature

1..1910  
/note="assembly\_fragment"  
2011..25284  
/note="assembly\_fragment"  
25385..30044  
/note="assembly\_fragment"  
30145..38883  
/note="assembly\_fragment"  
38984..62613  
/note="assembly\_fragment"  
62714..64667



```

/note="assembly_fragment"
64768. .66216
/note="assembly_fragment"
66317. 75523
/note="assembly_fragment"
75624. .78993
/note="assembly_fragment"
79094. .95809
/note="assembly_fragment"
95910. 98133
/note="assembly_fragment"
98234. .99926
/note="assembly_fragment"
100027. .122553
/note="assembly_fragment"
122654. 142576
/note="assembly_fragment"
142677. .157739
/note="assembly_fragment"
157840. .167044
/note="assembly_fragment"
167145. 170812
/note="assembly_fragment"
170913. .180254
/note="assembly_fragment
clone end:T7
vector_side:right"

ORIGIN
Query Match      68.1%; Score 18.4; DB 2; Length 180254;
Best Local Similarity 95.0%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      8 CAATTGAAGTATGATTCCT 27
|||||||
Db      93172 CAATTGAAGTATGATTCCT 93153

RESULT 37
AC110917
LOCUS      AC110917      193005 bp      DNA      linear      ROD 22-AUG-2003
DEFINITION Mus musculus, clone RP23-155I20, complete sequence.
ACCESSION  AC110917
VERSION     AC110917.11 GI:34099027
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 193005)
Birren,B., Nussbaum,C. and Lander,E.
Mus musculus, clone RP23-155I20
Unpublished

2 (bases 1 to 193005)
Birren,B., Linton,B., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Farrisa,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Lander,T., Lechoczky,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,N., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 193005)
Direct Submission
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (29-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 193005)
Direct Submission
Birren,B., Nussbaum,C. and Lander,E.
Submitted (22-AUG-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 22, 2003 this sequence version replaced gi:33300744.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22445
Center clone name: 155_I_20
-----
FEATURES
source
Location/Qualifiers
1. .193005
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-155I20"
/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature
1. .3637
/note="wgs_end_extension
clone end:SP6"
repeat_region
175. .769
/rpt_family="Lx"
repeat_region
770. .797
/rpt_family="(TAAA)n"
repeat_region
798. .1296
/rpt_family="Lx"
repeat_region
1297. .1338
/rpt_family="(CA)n"
repeat_region
1339. .1553
/rpt_family="Lx"
repeat_region
1963. .1995
/rpt_family="AT-rich"
repeat_region
2192. .2314
/rpt_family="RLTR13C"
repeat_region
2362. .6725

```



```

misc_feature      /rpt_family="Lx"
3638..3643
/note="clone boundary
clone end:SP6
site:ECORI"
repeat_region    /rpt_family="AT_rich"
6732..6781
complement(8778..10152)
/rpt_family="ORR1A-int"
10153..10175
/rpt_family="(CAAAA)n"
complement(10176..10400)
/rpt_family="ORR1A-int"
complement(10401..10736)
/rpt_family="ORR1A2"
11078..11157
/rpt_family="AT_rich"
11476..11506
/rpt_family="AT_rich"
11584..11754
/rpt_family="Lx5"
11887..11910
/rpt_family="AT_rich"
complement(13543..13815)
/rpt_family="ORR1A2"
complement(14225..14399)
/rpt_family="L1"
14427..14448
/rpt_family="AT_rich"
15058..15573
/rpt_family="L1"
15575..15754
/rpt_family="ORR1A2"
15755..17973
/rpt_family="L1"
17980..18892
/rpt_family="L1VL2"
18891..20698
/rpt_family="L1_RN"
20686..21519
/rpt_family="L1"
21951..23084
/rpt_family="L1"
23085..23122
/rpt_family="(CA)n"
23123..24060
/rpt_family="L1"
24062..28131
/rpt_family="L1"
28132..28354
/rpt_family="ETNerv2"
28734..28934
/rpt_family="L1_MW"
complement(28975..29087)
/rpt_family="URR1B"
complement(29347..29471)
/rpt_family="Lx9"
29563..29749
/rpt_family="Lx4"
29750..29773
/rpt_family="AT_rich"
complement(30096..30187)
/rpt_family="B3"
complement(30197..32010)
/rpt_family="Lx5"
complement(32153..34475)
/rpt_family="Lx2"
34476..34500
/rpt_family="AT_rich"
complement(34501..34658)
/rpt_family="B2_Mm2"
complement(34659..34808)
/rpt_family="Lx2"

complement(34807..35361)
/rpt_family="Lx2"
complement(35586..36658)
/rpt_family="Lx2"
36654..37726
/rpt_family="Lx2"
complement(37781..37846)
/rpt_family="L1_MW"
37856..37907
/rpt_family="(GAAAA)n"
37927..39279
/rpt_family="Lx2"
complement(39361..39633)
/rpt_family="MTD"
39771..40347
/rpt_family="MERVL"
40346..40666
/rpt_family="MERVL"
40695..41005
/rpt_family="Lx5"
complement(41009..41358)

Query Match      68.1%; Score 18.4; DB 10; Length 193005;
Best Local Similarity 95.0%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CAATTGAAGTTATGTATCT 27
|||||
Db 89552 CAATTGAAGTTATGTATCT 89571
|||||

RESULT 38
AC107233/c
LOCUS
DEFINITION Mus musculus clone RP23-151C4, WORKING DRAFT SEQUENCE, 11 unordered
pieces.
AC107233
AC107233.4 GI:28827989
HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 199432)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-151C4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 199432)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Katat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamazaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seanan,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
Submitted (17-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

```

3 (bases 1 to 199432)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
McEltrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramaeamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schnupack,R., Seaman,S., Sevrub,M.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 4, 2003 this sequence version replaced gi:28173220.
All repeats were identified using RepeatMasker:
Snit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19755
Center clone name: 151_C_4
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 197227 bases at least Q40
Consensus quality: 197905 bases at least Q30
Consensus quality: 198182 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 198432; sum-of-contigs
Quality coverage: 9.1 in Q20 bases; agarose-fp
Quality coverage: 8.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 27489: contig of 27489 bp in length
* 27490: gap of 100 bp
* 27590: contig of 1871 bp in length
* 29461: gap of 100 bp
* 29561: contig of 1551 bp in length
* 31111: contig of 100 bp
* 31212: contig of 85999 bp in length
* 117210: gap of 100 bp
* 117311: contig of 1784 bp in length
* 119094: gap of 100 bp
* 119194: gap of 100 bp
* 119195: contig of 3166 bp in length
* 122461: gap of 100 bp
* 122461: contig of 8378 bp in length
* 130839: gap of 100 bp
* 130839: contig of 11288 bp in length
* 142226: gap of 100 bp
* 142327: contig of 17089 bp in length
* 159415: contig of 17089 bp in length

FEATURES
             source
* 159416 159515: gap of 100 bp
* 159516 197685: contig of 38170 bp in length
* 197686 197785: gap of 100 bp
* 197786 199432: contig of 1647 bp in length.
             Location/Qualifiers
             1..199432
              /organism="Mus musculus"
              /mol_type="genomic DNA"
              /db_xref="taxon:10090"
              /clone="RP23-151C4"
              /clone_lib="RPCI-23 Female Mouse BAC"
             1..27489
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:left"
             27590..29460
              /note="assembly_fragment"
             29561..31111
              /note="assembly_fragment"
             31212..117210
              /note="assembly_fragment"
             117311..119094
              /note="assembly_fragment"
             119195..122360
              /note="assembly_fragment"
             122461..130838
              /note="assembly_fragment"
             130939..142226
              /note="assembly_fragment"
             142327..159415
              /note="assembly_fragment"
             159516..197685
              /note="assembly_fragment"
             197786..199432
              /note="assembly_fragment"
              clone_end:T7
              vector_side:right"

ORIGIN
Query Match      68.1%; Score 18.4; DB 2; Length 199432;
Best Local Similarity 95.0%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CAATTGAAGTTATGTATCTCT 27
Db 125847 CAATTGAAGTTATGTATCTCT 125828

RESULT 39
AC010172/c
LOCUS
DEFINITION
Homo sapiens chromosome 12 clone RP11-551G13, WORKING DRAFT
SEQUENCE 9 unordered pieces.
AC010172
AC010172.21 GI:20279305
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
VERSION
KEYWORDS
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 242301)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Hanks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carren,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyie,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

```

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, J., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulaeged, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheeshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshcari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 242301)  
Worley, K. C.

Direct Submission  
Submitted (15-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 242301)  
Worley, K. C.

Direct Submission  
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Apr 24, 2002 this sequence version replaced gi:14547712.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: HMHT  
Center clone name: RP11-551G13  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-primer Bodipy: 46% of reads  
Chemistry: Dye-terminator Big Dye: 54% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 234054 bases at least Q40  
Consensus quality: 236536 bases at least Q30  
Consensus quality: 238152 bases at least Q20  
Estimated insert size: 250093; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will be preserved.

1 3042: contig of 3042 bp in length  
3043 3142: gap of unknown length  
3143 6130: contig of 2988 bp in length  
6131 6230: gap of unknown length  
6231 14402: contig of 8172 bp in length  
14403 14502: gap of unknown length  
14503 31187: contig of 16685 bp in length  
31188 31287: gap of unknown length  
31288 55865: contig of 24578 bp in length  
55866 78599: contig of unknown length  
78599 78600: contig of 22634 bp in length  
78600 108718: contig of unknown length  
108719 108818: contig of 30019 bp in length  
108819 159929: contig of unknown length  
159930 160029: gap of unknown length  
160030 242301: contig of 82272 bp in length.

FEATURES  
Location/Qualifiers  
1..242301  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-551G13"

ORIGIN  
Query Match 68.1%; Score 18.4; DB 2; Length 242301;  
Best Local Similarity 95.0%; Pred. No. 6.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CAATTGAAGTTATGTATCTCT 27  
||||| |||||||  
Db 58390 CAATTGAGGTTATGTATCTCT 58371

RESULT 40  
SOLEU2  
LOCUS  
DEFINITION  
S.occidentalis LEU2 gene.  
ACCESSION  
X79823.1 GI:791136  
VERSION  
3-isopropylmalate dehydrogenase; LEU2 gene.  
KEYWORDS  
Debaryomyces occidentalis  
SOURCE  
Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Debaryomycetes; Saccharomycetales; Saccharomycetaceae; Debaryomycetes.

REFERENCE  
1  
AUTHORS  
Iserentant, D. and Verachtert, H.  
TITLE  
Cloning and sequencing of the LEU2 homologue gene of Schwannomyces occidentalis  
JOURNAL  
Yeast 11 (5), 467-473 (1995)  
MEDLINE  
95321019  
PUBMED  
7597851

REFERENCE  
2 (bases 1 to 1465)  
AUTHORS  
Iserentant, D. M. M.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (22-JUN-1994) D. M. M. Iserentant, University Leuven - Lab., Ind. Microbiol. & Biochem., Kardinaal Mercierlaan 92, 3001 Heverlee, BELGIUM

FEATURES  
Location/Qualifiers  
1..1465  
/organism="Debaryomyces occidentalis"  
/mol\_type="genomic DNA"  
/strain="ATCC26077/R91"  
/db\_xref="taxon:27300"  
144..1283  
/gene="LEU2"  
144..1283  
/gene="LEU2"  
/EC\_number="1.1.1.85"  
/codon\_start=1

/product="3-isopropylmalate dehydrogenase".  
/protein\_id="CAA56224.1"  
/db\_xref="GI:791137"  
/db\_xref="Swiss-Prot:P48012"  
/translation="MEISTAPNSSTYEPQIRNNQLIDTILNELLNVTNYTMGLVN  
LQNFILPURIQVLSTGIGDMTRINQVFPPTIDEIRINCILHDSLNKVSRYGIVEI  
IRVGMILPYFYKAFIRHEANLKYTRLNKFYSKNSKTIFSNLNINKGNYSREIDS  
IVSGSLELPKFKLILKRLHDAIESEKLNKQNFENFDNETSIITYYNTAIEVIDAF  
GGCGSSEDRKIDSOLFVFTPTGKILTELATKWPPPELQYGLSKRVVGIYELRNVPKS  
FDELEILIIIFSDHLEFLTIIESYVIRKNEPIKSLSVSDIIMHSLVNEKPLPVLSSL  
PSMEVNSWCNNEVITSCYLGVSSTSSKQDFLRLNVSKEGFS"

ORIGIN

Query Match 67.4%; Score 18.2; DB 8; Length 1465;  
Best Local Similarity 87.0%; Pred. No. 5.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 5 CCGCAATTGAAGTTATGATCCT 27  
Db 772 CCGCAATTGAAGTTATAGATGCT 794

Search completed: January 7, 2005, 11:33:30  
Job time : 938.789 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 205.579 Seconds  
(without alignments)  
689.440 Million cell updates/sec

Title: US-09-786-502A-7  
Perfect score: 27  
Sequence: 1 gggccgcgaattgaagtattatgcct 27

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 23Sep04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	27	3 AAA10270	Aaa10270 Human CD2
2	27	100.0	27	6 ABA92024	Abag2024 Human CD2
3	19	70.4	328	12 ADL67231	Adl67231 Human CD2
4	19	70.4	660	2 AAT96357	Aat96357 Human CD2
5	19	70.4	663	9 ADA02992	Ada02992 Human CD2
6	19	70.4	663	10 ADB72730	Adb72730 Human CD2
7	19	70.4	663	10 ADC85472	Adc85472 Human CD2
8	19	70.4	663	12 ADM74587	Adm74587 Human car
9	19	70.4	762	2 AAQ28837	Aaq28837 Sequence
10	19	70.4	1514	2 AAQ21167	Aaq21167 Human CD2
11	19	70.4	1514	2 AAT36257	Aat36257 Human CD2
12	19	70.4	1514	2 AAT14707	Aat14707 Human CD2
13	19	70.4	1514	2 AAV63445	Av63445 Human CD2
14	19	70.4	1514	2 AAV81203	Av81203 Human CD2
15	19	70.4	1514	3 AAS05081	Aas05081 Human cel
16	19	70.4	1514	3 AAZ29325	Aaz29325 Human CD2
17	19	70.4	1514	4 AAS03175	Aas03175 Human lym
18	19	70.4	1514	6 ABA99035	Abag99035 Human CD2
19	19	70.4	1514	10 ADD25537	Add25537 Binding d
20	19	70.4	1514	11 ADI31990	Adi31990 Human cDN
21	19	70.4	1514	12 ADO49343	Ado49343 Human CD2

C	22	19	70.4	1557	5 AAS89399	Aas89399 DNA encod
	23	19	70.4	1574	1 AAN90607	Aan90607 CD28 anti
	24	19	70.4	3803	6 AAL49546	Aal49546 Human CD2
	25	19	70.4	3804	8 ABV75351	Abv75351 Human CD2
	26	19	70.4	3804	9 ADA02991	Ada02991 Human CD2
	27	19	70.4	3804	10 ADB72729	Adb72729 Human CD2
	28	19	70.4	3804	10 ADC85471	Adc85471 Human CD2
	29	19	70.4	3804	12 ADM74586	Adm74586 Human car
	30	19	70.4	3804	12 ADN05855	Adn05855 Antipeori
	31	19	70.4	3806	6 AAD21976	Aad21976 Human CD2
	32	19	70.4	51365	9 ADA02990	Ada02990 Human CD2
	33	19	70.4	51365	10 ADB72728	Adb72728 Human CD2
	34	19	70.4	51365	10 ADC85470	Adc85470 Human CD2
	35	19	70.4	51365	12 ADM74585	Adm74585 Human car
	36	18.2	67.4	3692	3 AAA69169	Aaa69169 Bacteriop
	37	18.2	67.4	56506	3 AAA69168	Aaa69168 Bacteriop
	38	18	66.7	392	5 AAH94213	Aah94213 Human fce
	39	18	66.7	2000	12 ADJ41623	Adj41623 Plant cDN
	40	18	66.7	4093	4 ABL24354	AbL24354 Drosophil
	41	18	66.7	4878	10 ACF68174	Acf68174 Photorhab
	42	18	66.7	9265	4 ABL27568	AbL27568 Drosophil
	43	18	66.7	110000	2 AAV21209_05	Continuation (6 of
	44	18	66.7	110000	2 AAV21209_06	Continuation (7 of
	45	18	66.7	110000	10 ACF67367_08	Continuation (9 of

## ALIGNMENTS

## RESULT 1

AAA10270  
ID AAA10270 standard; DNA; 27 BP.

XX  
AC AAA10270;

XX  
DT 03-JUL-2000 (first entry)

XX  
DE Human CD28 fragment upstream PCR primer.

XX  
KW CD28; cytoplasmic domain; prostate-specific membrane antigen; PSMA;

XX  
KW J591 hybridoma; monoclonal antibody; single chain variable region; scFv;

XX  
KW single chain antibody; fusion receptor; immune response; prostate cancer;

XX  
KW PCR primer; ss.

XX  
OS Homo sapiens.

XX  
PN WC200014257-A1.

XX  
PD 16-MAR-2000.

XX  
PF 03-SBP-1999; 99WO-US020349.

XX  
PR 04-SEP-1998; 98US-0099138P.

XX  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.

XX  
PI Sadelain M, Bander NH, Gong M;

XX  
DR WPI; 2000-257002/22.

XX  
PT A fusion receptor composition having the structure:prostate-specific

XX  
PT membrane antigen-single chain variable fragment:optional

XX  
PT connector:cytoplasmic domain, useful for treatment of cancer.

XX  
XX Example 6; Page 15; 25pp; English.

XX  
XX The invention relates to a novel fusion receptor composition having the

XX  
XX structure: PSMA (prostate-specific membrane antigen)-scFv (single chain

XX  
XX variable fragment):optional connector:cytoplasmic domain, where the

XX  
XX fusion receptor is effective when expressed in a T-cell to promote a

XX  
XX cellular immune response to PSMA. The PSMA-scFv is a single-chain

XX  
XX antibody cloned from the V region genes of a hybridoma specific for PSMA,

XX  
XX such as J591. The optional connector is provided to give a spacing

CC between the PSMA-scfv and the cytoplasmic domain, such that both retain  
 CC substantial function. The cytoplasmic domain directs the function of the  
 CC fusion receptor and is generally the cytoplasmic domain of a molecule  
 CC which functions as a transducer of a mammalian immune response in the  
 CC presence of an MHC (major histocompatibility complex)-peptide complex or  
 CC costimulatory factor. Examples of cytoplasmic domains that may be  
 CC employed in the present invention include the T-cell receptor gamma-chain  
 CC cytoplasmic domain and the CD28 cytoplasmic domain. In a method of the  
 CC invention, an expression vector encoding the fusion receptor is  
 CC transduced into primary T-lymphocytes obtained from the patient to be  
 CC treated. The transduced lymphocytes are returned to the patient where  
 CC they secrete interleukin-2 (IL-2) and proliferate in response to PSMA-  
 CC positive cells. The resulting cytotoxic lymphocytes specifically lyse  
 CC cells expressing PSMA and can thus be used to target PSMA-positive tumour  
 CC cells. The fusion receptor promotes a cellular immune response to PSMA  
 CC and is useful for the treatment of prostate cancer and other cancers that  
 CC express PSMA. Sequences AAA10270-A10271 represent PCR primers used in an  
 CC exemplification of the present invention to amplify cDNA encoding the  
 CC cytoplasmic and transmembrane domains and part of the extracellular  
 CC domain of human CD28. This was used to construct a gene encoding a fusion  
 CC receptor comprising the PSMA-scfv and the CD28 fragment  
 XX  
 SQ Sequence 27 BP; 6 A; 6 C; 7 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 27; DB 3; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGGCGCAATTGAAGTTATGTATCCT 27  
 Db 1 GCGGCGCAATTGAAGTTATGTATCCT 27

RESULT 2  
 ABA92024  
 ID ABA92024 standard; DNA; 27 BP.  
 XX  
 AC ABA92024;  
 XX  
 DT 23-MAY-2002 (first entry)  
 XX  
 DE Human CD28 cDNA upstream PCR primer.  
 XX  
 KW CD28; T cell; receptor; human; antitumour; immunostimulant; cancer;  
 KW therapy; neuroblastoma; melanoma; sarcoma; small lung carcinoma;  
 KW brain tumour; disialoganglioside GD2; imaging; PCR; primer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002018783-A1.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 30-SEP-1997; 97US-00940544.  
 XX  
 PR 20-MAR-1997; 97WO-US004427.  
 XX

(SADE/) SADELAIN M.  
 (CHEU/) CHEUNG N V.  
 PA (KRAU/) KRAUSE A.  
 PA (GUOH/) GUO H.  
 XX  
 PI Sadelain M, Cheung NV, Krause A, Guo H;  
 DR WPI; 2002-239251/29.  
 XX  
 XX New fusion protein, useful for inducing host immune response, comprises  
 PT variable region of light chain of an antibody linked to variable region  
 PT of antibody, CD28 receptor signalling domain and transmembrane domain.  
 XX  
 XX Example 3; Page 4; 9pp; English.  
 PS  
 XX The present sequence is that of an upstream primer, used with the

CC downstream primer given in ABA92025, in the PCR amplification of a  
 CC segment of human T cell surface receptor CD28 cDNA. The amplified segment  
 CC encodes part of the extracellular domain of CD28, and the transmembrane  
 CC and the cytoplasmic domains. Plasmid pbsCD28 was used as template. The 5'  
 CC primer contains an NotI site, while the 3' primer contains a BamHI site.  
 CC and these sites in the PCR product were used to facilitate insertion into  
 CC retroviral vector SFG. A cell-surface molecule capable of CD28 signalling  
 CC in T cells interacting with disialoganglioside GD2 tumours was  
 CC constructed. This comprised the antigen-binding site of a GD2-specific  
 CC antibody and the transmembrane and signalling domains of the CD28  
 CC molecule. This is an example of fusion proteins of the invention  
 CC comprising a single chain Fv antibody linked to CD28 receptor. T cells  
 CC expressing such fusion proteins exhibit enhanced survival when  
 CC reintroduced to an in vivo environment. They can be used to induce an  
 CC immune response to cells, particularly tumour cells which express the  
 CC antigen to which the antibody is specific. Cells expressing a fusion  
 CC protein including an anti-GD2 scFv are useful for treatment of melanomas,  
 CC neuroblastomas, small lung carcinoma, sarcomas and brain tumours that  
 CC express GD2 as a surface antigen. Cells expressing the fusion proteins of  
 CC the invention can also be used for in vitro purging of stem cells or bone  
 CC marrow and for in vivo targeting of tumour cells and other antigen-  
 CC bearing cells for imaging  
 XX  
 SQ Sequence 27 BP; 6 A; 6 C; 7 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 27; DB 6; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGGCGCAATTGAAGTTATGTATCCT 27  
 Db 1 GCGGCGCAATTGAAGTTATGTATCCT 27

RESULT 3  
 ADL67231  
 ID ADL67231 standard; DNA; 328 BP.  
 XX  
 AC ADL67231;  
 XX  
 DT 20-MAY-2004 (first entry)  
 XX  
 DE Human CD28 costimulatory signalling element DNA.  
 XX  
 KW T cell receptor; TCR; CD3 zeta chain; co-stimulatory signalling region;  
 KW binding element; immunostimulant; therapy; cancer; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2004043401-A1.  
 XX  
 PD 04-MAR-2004.  
 XX  
 PF 28-MAY-2003; 2003US-00448256.  
 XX  
 PR 28-MAY-2002; 2002US-0383872P.  
 XX  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 PA  
 PI Sadelain M, Brentjens R, Maher J;  
 XX  
 XX WPI; 2004-225696/21.  
 XX  
 XX New nucleic acid polymer encoding a chimeric T cell receptor having a  
 PT zeta chain portion, useful for treating disorders where the immune  
 PT response needs to be induced, such as cancer.  
 XX  
 XX Example 7; SEQ ID NO 6; 25pp; English.  
 PS  
 XX The invention relates to a nucleic acid polymer encoding a chimeric T  
 CC cell receptor (TCR) which comprises human CD3 zeta chain intracellular  
 CC domain, a co-stimulatory signalling region and a binding element that  
 CC specifically interacts with a selected target. The methods and

CC compositions of the invention are useful for treating disorders where the  
 CC immune response needs to be induced, such as cancer. The present sequence  
 CC is human CD28 costimulatory signalling element DNA.

XX  
 XX Sequence 328 BP; 75 A; 94 C; 77 G; 82 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 19; DB 12; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27  
 |||||  
 Db 4 AATTGAAGTTATGATCCT 22

RESULT 4  
 AAT96357  
 ID AAT96357 standard; cDNA; 660 BP.

XX  
 AC AAT96357;  
 XX 08-APR-1998 (first entry)  
 XX  
 DE CD28 cDNA.  
 XX Screening; inhibitor; enhancer; binding; CD28; B7-1; ds.  
 XX Homo sapiens.

XX  
 FH Key Location/Qualifiers  
 FT CDS 1..660  
 FT /\*tag= a  
 FT /product= "CD28"

XX EP795554-A2.  
 PN 17-SEP-1997. 97BP-00301438.  
 XX  
 PF 04-MAR-1997; 97BP-00301438.  
 XX  
 PR 05-MAR-1996; 96JP-00047795.  
 PR 02-OCT-1996; 96JP-00262085.  
 XX (TAKE ) TAKEDA CHEM IND LTD.

XX Hida T, Hattori M, Kurokawa T, Nakanishi A;  
 XX WPI: 1997-450803/42.  
 DR P-PSDB; AAW38413.

XX New xanthene derivatives useful as immunomodulators - e.g. methyl 2-  
 PT (carboxymethylsulphinyl)-5,7-dichloro-3,8-dihydroxy-6-methyl-9-oxo-9H-  
 PT xanthene-1-carboxylate.

XX Disclosure; Page 81; 117pp; English.

XX The present sequence was used in the development of a novel method for  
 CC screening for compounds that inhibit or enhance binding of CD28 to B7-1  
 XX Sequence 660 BP; 164 A; 168 C; 150 G; 178 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 2; Length 660;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27  
 |||||  
 Db 339 AATTGAAGTTATGATCCT 357

RESULT 5  
 ADA02992  
 ID ADA02992 standard; cDNA; 663 BP.

XX  
 AC ADA02992;

XX 06-NOV-2003 (first entry)

XX Human CD28 carcinoma associated coding sequence, SEQ ID NO:1510.

XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 KW gene; ss.

XX Homo sapiens.

XX WO2003057146-A2.

XX 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.

XX 26-DEC-2001; 2001US-00035832.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI: 2003-587068/55.

XX New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.

XX Claim 1; SEQ ID NO 1510; 245pp; English.

CC The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 663 BP; 165 A; 168 C; 151 G; 179 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 9; Length 663;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27  
 |||||  
 Db 339 AATTGAAGTTATGATCCT 357

RESULT 6  
 ADB72730

ID ADB72730 standard; cDNA; 663 BP.

XX ADB72730;

XX 04-DEC-2003 (first entry)

XX DE Human CD28 cDNA.  
XX KW human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
XX KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
XX OS Homo sapiens.  
XX PN WO2003008583-A2.  
XX PD 30-JAN-2003.  
XX PF 26-DEC-2001; 2001WO-US051291.  
XX PR 02-MAR-2001; 2001US-00798586.  
XX PR 23-OCT-2001; 2001US-00004113.  
XX PR 08-NOV-2001; 2001US-00052482.  
XX PR 30-NOV-2001; 2001US-00997722.  
XX PR 20-DEC-2001; 2001US-00034650.  
XX PA (SAGR-) SAGRES DISCOVERY.  
XX PI Morris DW, Engelhard EK;  
XX DR WPI; 2003-239337/23.  
XX PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
XX PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
XX PS Claim 1; SEQ ID NO 58; 2304pp; English.  
XX CC The invention relates to a novel recombinant nucleic acid comprising a  
XX CC nucleotide sequence selected from any of the 660 sequences fully defined  
XX CC in the specification. A polynucleotide of the invention has cytostatic  
XX CC activity, and may have a use in gene therapy, or in a vaccine. The  
XX CC recombinant nucleic acids and polypeptides are useful for treating  
XX CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
XX CC sarcomas. The present sequence represents a human cDNA of the invention.  
XX SQ Sequence 663 BP; 165 A; 168 C; 151 G; 179 T; 0 U; 0 Other;  
XX  
XX Query Match 70.4%; Score 19; DB 10; Length 663;  
XX Best Local Similarity 100.0%; Pred. No. 42;  
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX RESULT 7  
XX ADC85472  
XX ID ADC85472 standard; DNA; 663 BP.  
XX XX  
XX AC ADC85472;  
XX DT 01-JAN-2004 (first entry)  
XX DE Human Cd28 coding sequence.  
XX KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;  
XX KW secreted; transmembrane; intracellular; ds.  
XX OS Homo sapiens.  
XX PN WO2003045230-A2.  
XX PD 05-JUN-2003.  
XX PF 02-DEC-2002; 2002WO-US038582.  
XX PR 30-NOV-2001; 2001US-00997722.  
XX XX

PA (SAGR-) SAGRES DISCOVERY.  
XX PI Morris DW, Engelhard EK;  
XX DR WPI; 2003-513603/48.  
XX PT New recombinant nucleic acid comprising a nucleotide sequence of any of  
XX PT the carcinoma-associated (CA) genes, useful for screening for drug  
XX PT candidates for diagnosing or treating carcinomas.  
XX PS Claim 1; SEQ ID NO 258; 983pp; English.  
XX CC The invention relates to a recombinant nucleic acid comprising a  
XX CC nucleotide sequence selected from any of the fully defined carcinoma-  
XX CC associated (CA) genes from the 50 tables given in the specification. The  
XX CC CA proteins are secreted, transmembrane or intracellular proteins. The  
XX CC recombinant nucleic acids are useful for screening for drug candidates  
XX CC for diagnosing or treating carcinomas. Sequences given in ADC85215-  
XX CC ADC85514 represent CA genes of the invention.  
XX SQ Sequence 663 BP; 165 A; 168 C; 151 G; 179 T; 0 U; 0 Other;  
XX  
XX Query Match 70.4%; Score 19; DB 10; Length 663;  
XX Best Local Similarity 100.0%; Pred. No. 42;  
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX RESULT 8  
XX ADM74587  
XX ID ADM74587 standard; DNA; 663 BP.  
XX XX  
XX AC ADM74587;  
XX DT 01-JUL-2004 (first entry)  
XX DE Human carcinoma associated (CA) nucleic acid #129.  
XX KW Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;  
XX KW carcinoma associated protein; CAP; carcinoma, leukaemia; lymphoma;  
XX KW cytostatic.  
XX OS Homo sapiens.  
XX PN US2004072154-A1.  
XX PD 15-APR-2004.  
XX PF 30-NOV-2001; 2001US-00997722.  
XX PR 22-DEC-2000; 2000US-00747377.  
XX PR 02-MAR-2001; 2001US-00798586.  
XX XX  
XX PA (MORR/) MORRIS D W.  
XX PA (ENGE/) ENGELHARD E K.  
XX PI Morris DW, Engelhard EK;  
XX XX  
XX DR WPI; 2004-328562/30.  
XX PT New carcinoma associated gene or protein, useful for preparing a  
XX PT composition for diagnosing or treating carcinoma e.g., leukemia or  
XX PT lymphoma.  
XX PS Claim 1; SEQ ID NO 258; 29pp; English.  
XX CC The invention relates to new recombinant nucleic acids. The invention  
XX CC also relates to a host cell comprising a recombinant nucleic acid or  
XX CC expression vector, an expression vector comprising a recombinant nucleic  
XX CC acid, a recombinant protein, a method of screening for drug candidates, a



CC method of screening for a bioactive agent capable of binding to a  
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a  
CC method of screening for a bioactive agent capable of modulating the  
CC activity of a CAP, a method of evaluating the effect of a candidate  
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting  
CC the activity of a CAP, a method of treating carcinomas, a method of  
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or  
CC propensity to carcinoma. A method of evaluating the effect of a candidate  
CC carcinoma drug comprises administering the drug to a patient, removing a  
CC cell sample from the patient and determining alterations in the  
CC expression or activation of a gene comprising the nucleotide sequence. A  
CC method of diagnosing carcinoma comprises determining the expression of  
CC one or more genes comprising the nucleic acid sequence in a first tissue  
CC type of a first individual and comparing the expression of the gene from  
CC a second normal tissue type from the first individual or a second  
CC unaffected individual, where a difference in the expression indicates  
CC that the first individual has carcinoma. A method of inhibiting the  
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating  
CC carcinomas comprises administering to a patient an inhibitor of CAP.  
CC Neutralising the effect of a CAP comprises contacting an agent specific  
CC for the CAP. The polypeptide specifically binds to the protein encoded by  
CC the nucleic acid. It comprises an antibody that specifically binds to the  
CC protein encoded by the nucleic acid. The nucleic acids are useful for  
CC preparing a composition for diagnosing or treating carcinoma e.g.,  
CC leukemia or lymphoma. This sequence represents a human carcinoma  
CC associated (CA) nucleic acid of the invention. Note: The sequence data  
CC for this patent did not form part of the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.

XX Sequence 663 BP; 165 A; 168 C; 151 G; 179 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 12; Length 663;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27  
DB 339 AATTGAAGTTATGATCCT 357

RESULT 9  
AAQ28837  
ID AAQ28837 standard; cDNA; 762 BP.

XX AC AAQ28837;

DT 25-MAR-2003 (revised)  
DT 25-FEB-1993 (first entry)

XX DE Sequence of the CD28 gene.

XX KW CD28 protein; T-cell immune response; Tp44; differentiation antigen;  
XX membrane protein; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX FT sig\_peptide 100..153  
XX FT mat\_peptide 154..762  
XX FT misc\_feature 538..552  
XX FT misc\_feature 556..570  
XX FT /tag= primer MIT  
XX FT /tag= d  
XX FT /label= primer C1T

XX PN WO9215671-A1.

XX PD 17-SEP-1992.

XX XX

PF 09-MAR-1992; 92WO-US001867.

PR 08-MAR-1991; 9IUS-0066809.

XX (CYTO-) CYTOMED INC.

XX PI Ko J, Ip SH;

XX DR WPI; 1992-331717/40.

XX DR P-PSDB; AAR27103.

XX PT Soluble proteins binding to B7 proteins and block antigen presenting  
XX cells - are useful in treating T-cell mediated immunosuppression diseases  
XX e.g. transplant rejection, auto immune diseases etc.

XX PS Disclosure; Fig 1; 43pp; English.

XX CC The cDNA sequence in AAQ28837 is derived from Aruffo A and Seed B, PNAS  
XX USA 84, 8573-77, 1987. The CD28 cDNA encodes a protein of 220 AAs  
XX (AAR27103). It contains a signal peptide from AAs 1-18 which is cleaved  
XX during the maturation of CD28 protein. The transmembrane region is AAs  
XX 153-79. For the production of soluble CD28 protein, the transmembrane  
XX region is deleted. The CD28 protein derived from P9-MIT comprises  
XX residues 19-157 and is not soluble; the CD28 protein derived from p9-CIT  
XX comprises residues 19-151 and is soluble. MIT and CIT are primers. T7  
XX primer is used as the 5' primer to pair with any 3' primer, eg, MIT and  
XX CIT, in a PCR reaction for the procurement of truncated CD28 gene which  
XX would direct the production of soluble CD28 protein. DNA encoding AAs 19-  
XX 151 of CD28 is claimed. (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 762 BP; 183 A; 198 C; 181 G; 200 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 2; Length 762;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27  
DB 438 AATTGAAGTTATGATCCT 456

RESULT 10  
AAQ21167  
ID AAQ21167 standard; DNA; 1514 BP.

XX AC AAQ21167;

DT 25-MAR-2003 (revised)  
DT 21-MAY-1992 (first entry)

XX DE Human CD28 antigen coding sequence.

XX KW cloning technique; cell surface antigen; immunodiagnosis; tumour; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX FT CDS 100..762  
XX FT /tag= a  
XX FT /product= "CD28"  
XX FT sig\_peptide 100..153  
XX FT /tag= b

XX PN WO9201049-A.

XX PD 23-JAN-1992.

XX PF 13-JUL-1990; 9OUS-00553759.

XX PR 13-JUL-1990; 9OUS-00553759.

XX PA (GEHO ) GEN HOSPITAL CORP.

XX XX



```

XX Example 3; Fig 7A-7B; 79pp; English.
XX A CDNA clone (AAT14707) codes for human CD28 antigen (AA891433). It was
CC obtd. by constructing an expression library in COS cells using cDNA
CC derived from human T-cell tumour line HPB-ALL and vector pIH3M (see also
CC AAT14705), and panning of the library using antibody-coated plates. This
CC immunoselection cloning method, developed to clone genes for cell surface
CC antigens of human lymphocytes (see also AAT14703-04 and AAT14706-26), has
CC general appln. Cell surface antigens are obtd. for diagnostic and
CC therapeutic use. (Updated on 25-MAR-2003 to correct PF field.)
XX Sequence 1514 BP; 404 A; 359 C; 337 G; 414 T; 0 U; 0 Other;
SQ Query Match 70.4%; Score 19; DB 2; Length 1514;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
Db 438 AATTGAAGTTATGATCCT 456

RESULT 13
AAV63445
ID AAV63445 standard; cDNA; 1514 BP.
XX AC AAV63445;
XX 25-MAR-2003 (revised)
DT 07-JUN-1999 (first entry)
XX Human CD28 cDNA.
DE Human CD28 cDNA.
XX CD28; cell surface antigen; human; lymphocyte; cloning; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 100..762
FT /*tag= a
FT sig_peptide 100..153
FT /*tag= b
FT mat_peptide 154..759
FT /*tag= c
XX US5830731-A.
XX 03-NOV-1998.
XX 21-MAY-1997; 97US-00861205.
XX 25-FEB-1988; 88US-00160416.
XX 13-JUL-1989; 89US-00379076.
XX 23-MAR-1990; 90US-00498809.
XX 13-JUL-1990; 90US-00553759.
XX 01-DEC-1992; 92US-00983647.
XX (GEO ) GEN HOSPITAL CORP.
XX Seed B, Aruffo A;
XX WPI; 1998-609251/51.
XX P-PSDB; AAW80442.
XX New cloning vector and poly-linker - based on existing sequences for
XX efficient cloning and expression of mammalian cDNA(s), especially human
XX lymphocyte antigenic sequences.
XX Example 3; Fig 7A-B; 75pp; English.
XX This nucleotide sequence comprises human CD28 cDNA. The cDNA was isolated
XX from HPB-ALL human T-cell tumour cells using a novel method for cloning

```

---

```

CC cDNAs from mammalian expression libraries. The method is based on
CC transient expression of an antigen in eukaryotic cells and physical
CC selection of cells expressing the antigen by adhesion to an antibody-
CC coated substrate. The method is useful for the isolation and molecular
CC cloning of any protein which can be expressed and transported to the cell
CC surface membrane of a eukaryotic cell. It has been used to clone genes
CC (see AAV63442-63) encoding cell surface antigens associated with
CC mammalian lymphocytes such as CD1-53, ICAM, LFA-3, FCRIa, FCRIb, TL1a
CC and Leu8 (see AAW80440-55). CD28 polypeptide (see AAW80442) has been
CC expressed in transfected COS cells. The purified genes and proteins are
CC useful for immunodiagnostic and immunotherapeutic applications, including
CC the diagnosis and treatment of immune-mediated infections, diseases, and
CC disorders of animals, including humans. (Updated on 25-MAR-2003 to
CC correct PR field.)
XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;
SQ Query Match 70.4%; Score 19; DB 2; Length 1514;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
Db 438 AATTGAAGTTATGATCCT 456

RESULT 14
AAV81203
ID AAV81203 standard; cDNA; 1514 BP.
XX AC AAV81203;
XX 10-MAY-1999 (first entry)
DT Human CD28 cDNA.
DE Human CD28 cDNA.
XX CD28; cell surface antigen; human; T cell antigen; T lymphocyte;
XX cDNA library; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 100..762
FT /*tag= a
FT sig_peptide 100..153
FT /*tag= b
FT mat_peptide 154..759
FT /*tag= c
XX US5849898-A.
XX 15-DEC-1998.
XX 07-JUN-1995; 95US-00485447.
XX 25-FEB-1988; 88US-00160416.
XX 13-JUL-1989; 89US-00379076.
XX 23-MAR-1990; 90US-00498809.
XX 13-JUL-1990; 90US-00553759.
XX 01-DEC-1992; 92US-00983647.
XX (GEO ) GEN HOSPITAL CORP.
XX Seed B, Oquendo C, Camerini D, Stamenkovic I, Stengelin S;
XX Amiot M, Lauffer L, Allen J, Simmons D, Aruffo A;
XX WPI; 1999-069813/06.
XX P-PSDB; AAW88451.
XX cDNA encoding human CD40 antigen - useful for cloning cDNA encoding cell
XX surface antigens, constructing cDNA libraries, expressing vectors for
XX expression in eukaryotic cells or their fragments.

```

PS Example 3; Fig 7A-B; 79pp; English.

XX This nucleotide sequence comprises human CD28 cDNA. The cDNA was isolated  
CC from HPB-ALL human T-cell tumour cells using a novel method for cloning  
CC cDNAs from mammalian expression libraries. The method is based on  
CC transient expression of an antigen in eukaryotic cells and physical  
CC selection of cells expressing the antigen by adhesion to an antibody-  
CC coated substrate. The method is useful for the isolation and molecular-  
CC cloning of any protein which can be expressed and transported to the cell  
CC surface membrane of a eukaryotic cell. It has been used to clone genes  
CC (see AAV81198-220) encoding cell surface antigens such as CD1a, CD1b,  
CC CD1c, CD2, CD6, CD7, CD13, CD14, CD16, CD19, CD20, CD22, CD26, CD27,  
CC CD28, CD31, CD32a, CD32b, CD33, CD34, CD36, CD37, CD38, CD39, CD40,  
CC CD43, CD44, CD53, ICAM, LFA-3, FCRIa, FCRIb, TLISA and Leu8 (see AAV86188  
CC -62, AAV89151-52, and AAV88451). CD40 cDNA (see AAV81198) is specifically  
CC claimed. CD28 polypeptide (see AAV88451) has been expressed in  
CC transfected COS cells

XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 2; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTATCCT 27  
Db 438 AATTGAAGTTATGTATCCT 456

RESULT 15  
AAA50581  
ID AAA50581 standard; cDNA; 1514 BP.

AC AAA50581;

XX 19-DEC-2000 (first entry)

XX Human cell surface antigen CD28 cDNA.

XX CD28; cell surface antigen; human; immunoselection; panning;  
KW immunodiagnosis; diagnosis; immunotherapy; gene therapy; immune disorder;  
KW infection; asthma; immune-complex disease; amyloidosis;  
KW multiple sclerosis; parasitic disease; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 100..762  
FT sig\_peptide /\*tag= a  
FT FT 100..153  
FT mat\_peptide /\*tag= b  
FT FT 154..759  
FT /\*tag= c

XX US6111093-A.

XX 29-AUG-2000.

XX 28-OCT-1998; 98US-00181612.

XX 25-FEB-1988; 88US-00160416.

XX 13-JUL-1989; 89US-00379076.

XX 23-MAR-1990; 90US-00498809.

XX 13-JUL-1990; 90US-00553759.

XX 01-DEC-1992; 92US-00983647.

XX (GEO) GEN HOSPITAL CORP.

XX Stamenkovic I, Seed B;

XX WPI; 2000-586382/55.

XX P-PSDB; AAY96128.

PT Isolated nucleic acid molecule encoding the CD19 cell surface antigen,  
PT useful for immunodiagnosis and immunotherapy of immune-mediated  
PT infections or disorders, e.g. asthma, immune-complex disease, parasitic  
PT diseases.

XX Example 3; Fig 7A-B; 75pp; English.

XX The present sequence is that of cDNA encoding human cell surface antigen  
CC (CSA) CD28 (see AAY96128). The cDNA was derived from a human T-cell  
CC tumour line HPB-ALL cDNA library using a new method for cloning CSA  
CC cDNAs. The method is based upon transient expression of CSA in eukaryotic  
CC cells and physical selection of cells expressing the antigen by adhesion  
CC to [panning on] an antibody-coated substrate such as a culture dish. The  
CC predicted amino acid sequence of CD28 suggests an integral membrane  
CC protein with a single membrane-spanning hydrophobic domain terminating in  
CC a 41-amino acid cytoplasmic domain. CSA nucleic acids isolated by the  
CC method of the invention, and the proteins they encode, are useful for  
CC immunodiagnostic and immunotherapeutic applications, including the  
CC diagnosis and treatment of immune-mediated infections, diseases, and  
CC disorders in animals, including humans. These disorders include asthma,  
CC immune-complex disease, amyloidosis, parasitic diseases or multiple  
CC sclerosis

XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 3; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTATCCT 27  
Db 438 AATTGAAGTTATGTATCCT 456

RESULT 16

AAZ29325  
ID AAZ29325 standard; cDNA; 1514 BP.

XX AAZ29325;

XX 29-FEB-2000 (first entry)

XX Human CD28 receptor cDNA.

XX Human CD28; B7 molecule; immune response; cell surface receptor;  
KW Major histocompatibility complex; MHC classII; proton motor force;  
KW mitochondrial membrane potential; mitochondrial metabolism; cancer;  
KW autoimmune disease; neurodegenerative disorder; ds.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 100..762  
FT /\*tag= a  
FT /\*product= "Human CD28"

XX WO9953953-A2.

XX 28-OCT-1999.

XX 30-MAR-1999; 99WO-US006874.

XX 17-APR-1998; 98US-0082250P.

XX 29-JUL-1998; 98US-0094519P.

XX 24-SEP-1998; 98US-0101580P.

XX (UYVE-) UNIV VERMONT.

XX Newell MK;

XX WPI; 2000-096773/08.

XX P-PSDB; AAY44294.

PT Use of cell surface and membrane characteristics for developing products  
 PT for treating cancers, autoimmune diseases or neurodegenerative diseases.  
 XX  
 PS Disclosure; Page 120-121; 123pp; English.

XX The present sequence encodes human CD28 receptor. CD28 is a homodimeric  
 CC glycoprotein involved in a secondary signalling pathway in the activation  
 CC of T-cell proliferation. B7 on nerve cells can interact with CD28 on the  
 CC immune cell leading to immune cell activation. The regulation of cell  
 CC surface expression of MHC classII and co-stimulatory molecule B7 can be  
 CC manipulated by regulating the intracellular dissipation of proton motor  
 CC force which can be assessed in terms of mitochondrial membrane potential.  
 CC These methods can be used for regulating cell growth and division to  
 CC control disease processes by manipulating mitochondrial metabolism and  
 CC the expression of cell surface immune proteins. They can be used for  
 CC treating diseases associated with excessive cellular division, aberrant  
 CC differentiation, and premature cellular death, e.g. cancers, autoimmune  
 CC diseases, neurodegenerative disorders etc

SQ Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 3; Length 1514;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTCCT 27  
 |||||  
 Db 438 AATTGAAGTTATGTCCT 456

RESULT 17  
 AAS03175  
 ID AAS03175 standard; cDNA; 1514 BP.

XX AAS03175;

AC AAS03175;

DT 29-AUG-2001 (first entry)

DE Human lymphocyte cell surface antigen CD28 cDNA sequence.

XX Human; lymphocyte cell surface antigen; immune-mediated disease; CD28;  
 KW infection; immune deficiency disorder; hypersensitivity; inflammation;  
 KW systemic lupus erythematosus; platelet disorder; rheumatoid arthritis;  
 KW transplant rejection; asthma; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FT CDS 100..762

FT /\*tag= a /product= "CD28 antigen"  
 FT sig\_peptide 100..153

FT /\*tag= b

FT mat\_peptide 154..759

FT /\*tag= c

XX US6218525-B1.

PN 17-APR-2001.

XX 01-DEC-1992; 92US-00983647.

XX 25-FEB-1988; 88US-00160416.

PR 13-JUL-1989; 89US-00379076.

PR 13-JUL-1990; 90US-00553759.

XX (GEO ) GEN HOSPITAL CORP.

XX Seed B, Aruffo A, Simmons D;

XX WPI; 2001-289848/30.

DR P-PSDB; AAU02437.

PT

PT

XX

XX

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

PT

PT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT

PT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT

PT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT

PT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT

PT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT

PT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT

PT Inhibiting an immune response in a subject for treating autoimmune  
PT disease such as psoriasis, diabetes mellitus, multiple sclerosis,  
PT rheumatoid arthritis, polymyositis, by administering anti-CD28 antibody.  
XX  
XX Disclosure; Page 28; 37pp; English.  
XX  
XX The sequence represents the human CD28 gene. The invention relates to a  
CC novel method for inhibiting an immune response by administering an anti-  
CC CD28 antibody. The method of the invention has immunosuppressive,  
CC antipsoriatic, antidiabetic, antiarthritic, antirheumatic,  
CC neuroprotective, dermatological, vasotropic, antiinflammatory,  
CC antianemic, hepatotropic, antiulcer, antibacterial and ophthalmological  
CC activity. The method works to inhibit immune response by reversing or  
CC blocking T cell activation. The method is useful for inhibiting an immune  
CC response in a subject susceptible to graft-versus-host disease (GVHD),  
CC marrow transplant rejection, organ transplant rejection or tissue  
CC transplant rejection, or having autoimmune disease including psoriasis,  
CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, systemic  
CC lupus erythematosus, systemic sclerosis, dermatomyositis, polymyositis,  
CC Sjogren's syndrome, polyarteritis nodosa or vasculitis. The method is  
CC also useful for treating sepsis, and other autoimmune diseases including  
CC autoimmune hepatitis, autoimmune haemolytic anaemia, Behcet's disease,  
CC myasthenia gravis, cirrhosis, uveitis, ulcerative colitis, and vitiligo  
XX  
XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 19; DB 6; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 AATTGAAGTTATGTCCT 27  
DB 438 AATTGAAGTTATGTCCT 456  
RESULT 19  
ADD25537  
ID ADD25537 standard; DNA; 1514 BP.  
XX AC ADD25537;  
XX  
XX 15-JAN-2004 (first entry)  
XX  
XX Binding domain-immunoglobulin fusion protein-associated DNA #53.  
DE  
XX db; Binding domain; immunoglobulin; fusion protein; cytostatic;  
XX antiarthritic; immunosuppressive; antidiabetic; antithyroid;  
XX neuroprotective; hinge region; immunoglobulin heavy chain;  
XX CH2 constant region; CH3 constant region; IGG1;  
XX antibody dependent cell-mediated cytotoxicity; ADCC; complement fixation;  
XX malignant condition; B-cell disorder; melanoma; carcinoma; sarcoma;  
XX rheumatoid arthritis; myasthenia gravis; Grave's disease;  
XX type I diabetes mellitus; multiple sclerosis; autoimmune disease.  
XX  
XX Unidentified.  
XX  
XX US2003118592-A1.  
XX  
XX 26-JUN-2003.  
XX  
XX 25-JUL-2002; 2002US-00207655.  
XX  
XX 17-JAN-2001; 2001US-0367358P.  
XX 17-JAN-2002; 2002US-00053530.  
XX 03-JUN-2002; 2002US-0385691P.  
XX  
XX (GENE-) GENE-CRAFT INC.  
XX  
XX Ledbetter JA, Hayden-Ledbetter MS, Thompson PA;  
XX WPI; 2003-801317/75.  
XX  
XX New binding domain-immunoglobulin fusion protein, useful for treating a

PT subject having or suspected of having a malignant condition or a B-cell  
PT disorder, e.g. melanoma, Grave's disease or autoimmune disease.  
XX  
XX Disclosure; SEQ ID NO 98; 157pp; English.  
XX  
XX Unidentified  
XX  
XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 19; DB 10; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 AATTGAAGTTATGTCCT 27  
DB 438 AATTGAAGTTATGTCCT 456  
RESULT 20  
ADD131990  
ID ADD131990 standard; cDNA; 1514 BP.  
XX AC ADD131990;  
XX  
XX 17-JUN-2004 (first entry)  
XX  
XX Human cDNA #1316.  
XX  
XX Human; gene; ss; immunological response; immunopathological condition;  
XX Crohn's disease; asthma; ulcerative colitis; hyperesinophilia;  
XX irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;  
XX acute monocytic leukaemia; antinflammatory; antiasthmatic; antiulcer;  
XX osteopathic; antiarthritic; antirheumatic; cytostatic.  
XX  
XX Homo sapiens.  
XX  
XX US6607879-B1.  
XX  
XX 19-AUG-2003.  
XX  
XX 09-FEB-1998; 98US-00023655.  
XX  
XX 09-FEB-1998; 98US-00023655.  
XX  
XX (INCY-) INCYTE CORP.  
XX  
XX Cocks BG, Stuart SG, Seilhamer JJ;  
XX WPI; 2003-895307/82.  
XX  
XX A composition comprising a plurality of cDNAs, useful for detecting  
PT altered expression of genes in an immunological response or for  
PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma  
PT or osteoarthritis.  
XX  
XX Claim 1; SEQ ID NO 1316; 50pp; English.  
XX  
XX The invention relates to a composition comprising a plurality of cDNAs  
CC for detecting the altered expression of genes in an immunological  
CC response. The invention also relates to a method of diagnosing or  
CC monitoring the treatment of an immunopathological condition in a sample,  
CC comprising obtaining nucleic acids from a sample, contacting the nucleic  
CC acids of the sample with an array comprising the plurality of cDNAs under  
CC conditions to form one or more hybridisation complexes, detecting the  
CC hybridisation complexes and comparing the levels of the detected  
CC hybridisation complexes with the level of hybridisation complexes  
CC detected in a non-diseased sample, where an altered level of the detected  
CC hybridisation complexes correlates with the presence of an  
CC immunopathological condition. Also disclosed are an expression profile  
CC comprising a microarray and a plurality of detectable complexes and a  
CC method for identifying a plurality of polynucleotide probes. The cDNAs  
CC are useful as hybridisable array elements in a microarray for monitoring  
CC the expression of target polynucleotides. The microarray can be used in

CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,  
 CC ulcerative colitis, hyperesinophilia, irritable bowel syndrome,  
 CC osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in  
 CC identifying agents for the treatment of the diseases. The microarray may  
 CC also be used in drug discovery and development, toxicological and  
 CC carcinogenicity studies, forensics or pharmacogenomics. The composition  
 CC may also be used in purification of a subpopulation of mRNAs, cDNAs or  
 CC genomic fragments. This sequence represents a human cDNA of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification but was obtained in electronic format directly  
 CC from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).

SQ Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 11; Length 1514;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27  
 |||||  
 Db 438 AATTGAAGTTATGATCCT 456

## RESULT 21

ADO49343  
 ID ADO49343 standard; cDNA; 1514 BP.

XX  
 AC ADO49343;

DT 15-JUL-2004 (first entry)

DE Human CD28 antigen cDNA.

XX cell surface antigen; immune-mediated disorder; asthma;  
 KW rheumatoid arthritis; multiple sclerosis; vasculitis; inflammation; ss;  
 KW gene; human.

OS Homo sapiens.

XX US2004072283-A1.

XX 15-APR-2004.

PF 17-APR-2001; 2001US-00836544.

XX 25-FEB-1988; 88US-00160416.

PR 13-JUL-1989; 89US-00379076.

PR 23-MAR-1990; 90US-00498809.

PR 13-JUL-1990; 90US-00553759.

PR 01-DEC-1992; 92US-00983647.

XX (SEED/) SEED B.

PA (ALLE/) ALLEN J.

PA (ARUF/) ARUFFO A.

PA (CAME/) CAMERINI D.

PA (LAUF/) LAUFFER L.

PA (OQUE/) OQUENDO C.

PA (SIMM/) SIMMONS D.

PA (STAM/) STAMENKOVIC I.

PA (STEN/) STENGELIN S.

PA (AMIO/) AMIOT M.

XX Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo C;

PI Simmons D, Stamenkovic I, Stengelin S, Amiot M;

XX WPI; 2004-328571/30.

DR P-PSDB; ADO49344.

XX New cloning cDNA segments encoding cell surface antigens of human

PS Example 3; Fig 7; 75pp; English.

XX The invention relates to a cloned cDNA segment encoding a cell surface  
 CC antigen selected from CD1a, CD1b, CD1c, CD2, CD6, CD7, CD13, CD14, CD16,  
 CC CD19, CD20, CD22, CD26, CD31, CD32, CD33, CD34, CD36, CD37, CD38,  
 CC CD39, CD40, CD43, and their functional derivatives. The cell  
 CC surface antigens of human lymphocytes prepared from the cDNAs are useful  
 CC in diagnostic and therapeutic utility in immune-mediated disorders  
 CC (asthma, rheumatoid arthritis, multiple sclerosis, vasculitis and  
 CC inflammation) and infections in mammals, including humans. The present  
 CC sequence represents a human cell surface antigen cDNA.

SQ Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 70.4%; Score 19; DB 12; Length 1514;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27  
 |||||  
 Db 438 AATTGAAGTTATGATCCT 456

## RESULT 22

AAS89399/c  
 ID AAS89399 standard; cDNA; 1557 BP.

XX  
 AC AAS89399;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #25203.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG25212.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

PS Claim 1; SEQ ID NO 25203; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 SQ Sequence 1557 BP; 392 A; 398 C; 361 G; 406 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 19; DB 5; Length 1557;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 AATTGAAGTTATGTCCT 27  
 Db 193 AATTGAAGTTATGTCCT 175  
 RESULT 23  
 AAN90607  
 ID AAN90607 standard; cDNA; 1574 BP.  
 XX  
 AC AAN90607;  
 XX  
 DT 09-SEP-2004 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 31-OCT-2002 (revised)  
 DT 20-DEC-1989 (first entry)  
 XX  
 DE CD28 antigen cDNA.  
 XX  
 KW Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;  
 KW HIV box; immunoselection; immune deficiency diseases; vasculitis;  
 KW systemic lupus erythematosus; rheumatoid arthritis; neoplasms; ss.  
 XX  
 OS Homo sapiens.  
 OS Unidentified.  
 XX  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 100..819  
 FT /\*tag= a  
 FT sig\_peptide 100..274  
 FT /\*tag= b  
 XX  
 PN EP330191-A.  
 XX  
 PD 30-AUG-1989.  
 XX  
 XX  
 PF 23-FEB-1989; 89EP-00103127.  
 XX  
 PF 25-FEB-1988; 88US-00160416.  
 PR  
 XX (GEMO ) GEN HOSPITAL CORP.  
 PA  
 XX Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP;  
 PI Simmons D, Stamenkovi I;  
 PI  
 XX WPI; 1989-250302/35.  
 DR P-PSDB; AAP91324.  
 DR  
 XX Rapid immuno:selection cloning - used to clone genes encoding cell  
 PT surface antigens associated with mammalian T lymphocytes.  
 XX  
 XX Disclosure; Fig 7; 69pp; English.  
 PS  
 XX This is used for cloning into a vector which transforms COS cells. The  
 CC vectors can be used to isolate any protein by immunoselection, and clones  
 CC are easy to manipulate. (Updated on 31-OCT-2002 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC

CC Revised record issued on 09-SEP-2004 : Correction to keywords  
 XX  
 SQ Sequence 1574 BP; 423 A; 375 C; 347 G; 429 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 19; DB 1; Length 1574;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 AATTGAAGTTATGTCCT 27  
 Db 498 AATTGAAGTTATGTCCT 516  
 RESULT 24  
 AAL49546  
 ID AAL49546 standard; DNA; 3803 BP.  
 XX  
 AC AAL49546;  
 XX  
 DT 27-NOV-2002 (first entry)  
 XX  
 DE Human CD28 coding sequence.  
 XX  
 KW Human; CD28; immune response; signaling; diabetes; autoimmune disease;  
 KW immunosuppressive; dermatological; antiinflammatory; antithyroid;  
 KW antirheumatic; antiarthritic; neutropic; allergy; rheumatoid arthritis;  
 KW systemic lupus erythematosus; myasthenia gravis; autoimmune thyroiditis;  
 KW vitiligo; alopecia; inflammatory bowel disease; Addison's disease;  
 KW Graves disease; haemolytic anaemia; Sjogren's syndrome; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 222..884  
 FT /\*tag= a  
 FT /product= "CD28"  
 FT /transl\_except= (pos:867..872,aa:Ala)  
 XX  
 PN WO200266059-A2.  
 XX  
 PD 29-AUG-2002.  
 XX  
 PF 15-FEB-2002; 2002WO-US004772.  
 XX  
 XX 16-FEB-2001; 2001US-0269756P.  
 PR  
 XX (GEMY ) GENETICS INST INC.  
 PA  
 XX O'hara RM, Nagelin AM;  
 PI  
 XX WPI; 2002-674899/72.  
 DR P-PSDB; AA019101.  
 DR  
 XX Downmodulating immune responses by blocking CD28-mediated signaling,  
 PT useful for preventing and/or treating autoimmune disorders such as  
 PT diabetes, allergic reactions, graft versus host disease, systemic lupus  
 PT erythematosus.  
 XX  
 XX Disclosure; Page 60-61; 61pp; English.  
 PS  
 XX The present invention relates to a method of therapeutically  
 CC downmodulating an autoimmune response or an ongoing autoimmune response,  
 CC comprising administering an antigen binding portion of an anti-CD28  
 CC antibody that blocks signaling via CD28 to the subject so that an  
 CC autoimmune response or an ongoing autoimmune response in the subject is  
 CC downmodulated. The methods are useful in therapeutically and  
 CC prophylactically downmodulating the immune response in subjects having  
 CC autoimmune disorders such as diabetes, allergy and allergic reactions,  
 CC transplantation rejection, graft versus host disease, systemic lupus  
 CC erythematosus, rheumatoid arthritis, myasthenia gravis, autoimmune  
 CC thyroiditis, vitiligo, alopecia, inflammatory bowel disease, Addison's  
 CC disease, Graves disease, haemolytic anaemia and Sjogren's syndrome. The  
 CC present sequence is the human CD28 coding sequence  
 CC



```
XX SQ Sequence 3803 BP; 1076 A; 766 C; 880 G; 1081 T; 0 U; 0 Other;
Query Match 70.4%; Score 19; DB 6; Length 3803;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTATCCT 27
DB 561 AATTGAAGTTATGTATCCT 579

RESULT 26
ADA02991
ID ADA02991 standard; cDNA; 3804 BP.
XX AC ADA02991;
XX DT 06-NOV-2003 (first entry)
XX DE Human CD28 carcinoma associated cDNA, SEQ ID NO:1509.
XX KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX KW gene; ss.
XX OS Homo sapiens.
XX PN WO2003057146-A2.
XX PD 17-JUL-2003.
XX PF 26-DEC-2002; 2002WO-US041414.
XX PR 26-DEC-2001; 2001US-00035832.
XX PA (SAGR-) SAGRES DISCOVERY.
XX PI Morris DW;
XX DR WPI; 2003-587068/55.
XX PT New recombinant nucleic acid encoding carcinoma associated protein,
XX PT useful for preparing compositions for treating carcinomas.
XX PS Claim 1; SEQ ID NO 1509; 245pp; English.
XX CC The invention relates to recombinant carcinoma associated (CA) nucleic
XX CC acid sequences from mouse and human (ADA01482-ADA03094), and to
XX CC recombinant carcinoma associated proteins (CAP) encoded by them. The
XX CC invention also encompasses expression vectors and host cells comprising a
XX CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
XX CC binds to the protein, and a biochip comprising CA nucleic acid or
XX CC fragments thereof. The sequences of the invention were identified using
XX CC oncogenic retroviruses, which insert into the genome of the host organism
XX CC at random. Many of these do not carry transduced host oncogenes or
XX CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
XX CC direct consequence of the effects of proviral integration into host
XX CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
XX CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
XX CC leukaemia) or a propensity to carcinoma by determination of the sequence
XX CC of a CA gene, or by determination of CA gene expression in particular
XX CC tissues. CA nucleic acids, proteins and antibodies are also useful as
XX CC therapeutic agents and in screening and evaluating drug candidates. The
XX CC present sequence represents a specifically claimed human CA nucleic acid
XX CC sequence of the invention. Note: The complete sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;
Query Match 70.4%; Score 19; DB 9; Length 3804;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTATCCT 27
DB 561 AATTGAAGTTATGTATCCT 579

us-09-786-502a-7.rng
XX SQ Sequence 3803 BP; 1076 A; 766 C; 880 G; 1081 T; 0 U; 0 Other;
Query Match 70.4%; Score 19; DB 6; Length 3803;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTATCCT 27
DB 560 AATTGAAGTTATGTATCCT 578

RESULT 25
ABV75351
ID ABV75351 standard; DNA; 3804 BP.
XX AC ABV75351;
XX DT 07-MAR-2003 (first entry)
XX DE Human CD28 protein encoding DNA.
XX KW T-cell; cell proliferation; cell survival; cytostatic; antiallergic;
XX KW immunostimulant; immunosuppressive; gene therapy; CD28; human; gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 223..885
XX FT /*tag= a
XX FT /product= "CD28"
XX PN WO200290518-A2.
XX PD 14-NOV-2002.
XX PF 10-MAY-2002; 2002WO-US014843.
XX PR 10-MAY-2001; 2001US-0290097P.
XX PA (UNIW ) UNIV WASHINGTON.
XX PI Green JM, Shaw AS;
XX WPI; 2003-111968/10.
XX P-PSDB; ABB82721.
XX Selectively modulating T-cell survival or proliferation, useful for
XX treating autoimmune diseases, comprises contacting T-cells expressing
XX CD28 protein with agents that modulate activity of CD28 survival or
XX proliferation motif.
XX Disclosure; Page 115-117; 119pp; English.
XX The invention relates to selectively modulating T-cell survival relative
XX to T-cell proliferation, or T-cell proliferation relative to T-cell
XX survival. The method involves contacting a T-cell expressing a CD28
XX protein with an agent that selectively modulates the activity of a CD28
XX survival or proliferation motif relative to the activity of a CD28
XX proliferation or survival motif, respectively, to selectively modulate
XX survival or proliferation of the T-cell. The method is useful in
XX selectively modulating survival or proliferation of T-cells in a subject.
XX The method may also be used to prevent or treat disorders associated with
XX aberrant cell proliferation or survival, e.g. cancer, transplant-
XX associated disorders, allergic diseases, graft-versus-host disease,
XX lymphoproliferative disorders or autoimmune diseases. The present
XX sequence represents a DNA encoding the human CD28 protein
XX SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;
Query Match 70.4%; Score 19; DB 8; Length 3804;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



PT composition for diagnosing or treating carcinoma e.g., leukemia or  
 PT lymphoma.  
 PS Claim 1; SEQ ID NO 257; 29pp; English.  
 XX  
 CC The invention relates to new recombinant nucleic acids. The invention  
 CC also relates to a host cell comprising a recombinant nucleic acid or  
 CC expression vector, an expression vector comprising a recombinant nucleic  
 CC acid, a recombinant protein, a method of screening for drug candidates, a  
 CC method of screening for a bioactive agent capable of binding to a  
 CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a  
 CC method of screening for a bioactive agent capable of modulating the  
 CC activity of a CAP, a method of evaluating the effect of a candidate  
 CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting  
 CC the activity of a CAP, a method of treating carcinomas, a method of  
 CC neutralising the effect of a CAP and a method of diagnosing carcinoma or  
 CC propensity to carcinoma. A method of evaluating the effect of a candidate  
 CC carcinoma drug comprises administering the drug to a patient, removing a  
 CC cell sample from the patient and determining alterations in the  
 CC expression or activation of a gene comprising the nucleotide sequence. A  
 CC method of diagnosing carcinoma comprises determining the expression of  
 CC one or more genes comprising the nucleic acid sequence in a first tissue  
 CC type of a first individual and comparing the expression of the gene from  
 CC a second normal tissue type from the first individual or a second  
 CC unaffected individual, where a difference in the expression indicates  
 CC that the first individual has carcinoma. A method of inhibiting the  
 CC activity of a CAP comprises binding an inhibitor to the CAP. Treating  
 CC carcinomas comprises administering to a patient an inhibitor of CAP.  
 CC Neutralising the effect of a CAP comprises contacting an agent specific  
 CC for the CAP. The polypeptide specifically binds to the protein encoded by  
 CC the nucleic acid. It comprises an antibody that specifically binds to the  
 CC protein encoded by the nucleic acid. The nucleic acids are useful for  
 CC preparing a composition for diagnosing or treating carcinoma e.g.,  
 CC leukaemia or lymphoma. This sequence represents a human carcinoma  
 CC associated (CA) nucleic acid of the invention. Note: The sequence data  
 CC for this patent did not form part of the printed specification but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html.  
 XX  
 SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 19; DB 12; Length 3804;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 AATTGAAGTTATGTCCT 27  
 DB 561 AATTGAAGTTATGTCCT 579  
 RESULT 31  
 AAD21976  
 ID AAD21976 standard; DNA; 3806 BP.  
 XX  
 AC AAD21976;  
 XX  
 DT 12-FEB-2002 (first entry)  
 XX  
 DE Human CD28 DNA.  
 XX  
 KW Human; cytotoxic T lymphocyte antigen-4; CTLA-4; CD28; AIDS; therapy;  
 KW Acquired Immune Deficiency Syndrome; neuroprotective; dermatological;  
 KW immune response; organ transplantation; autoimmune disease; allergy; SLE;  
 KW systemic lupus erythematosus; multiple sclerosis; tumour vaccination;  
 KW immunodeficiency disease; Digeorge Syndrome; cancer; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 223..885  
 FT /\*tag= a  
 FT /product= "Human CD28 protein"  
 XX  
 PN WO200179300-A1.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 12-APR-2001; 2001WO-US012275.  
 XX  
 PR 12-APR-2000; 2000US-0196851P.  
 XX  
 PA (GEMY ) GENETICS INST INC.  
 XX  
 PI Bluestone JA, Collins M, Whitters M, Griffin M, Kranz D;  
 XX  
 DR WPI: 2002-017603/02.  
 DR P-PSDB; AAE13721.  
 XX  
 PT New construct for downmodulating immune response in a subject, has  
 PT exposed surface attached with antigen-binding portion of antibody that  
 PT binds to cytotoxic T lymphocyte antigen-4 and major histocompatibility  
 PT molecule.  
 XX  
 PA (GETH ) GENENTECH INC.

XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
 PI Wu TD;  
 XX  
 DR WPI: 2004-305105/28.  
 DR P-PSDB; ADN05856.  
 XX  
 PT New PRO nucleic acid or polypeptide, useful for preparing a  
 PT pharmaceutical composition for diagnosing or treating psoriasis in a  
 PT mammal.  
 XX  
 PS Claim 1; SEQ ID NO 2250; 3069pp; English.  
 XX  
 CC The invention relates to novel polynucleotide and polypeptides for  
 CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition  
 CC for diagnosing or treating psoriasis in a mammal. This sequence  
 CC corresponds to one of the polynucleotides of the invention.  
 XX  
 SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 19; DB 12; Length 3804;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 AATTGAAGTTATGTCCT 27  
 DB 561 AATTGAAGTTATGTCCT 579  
 RESULT 31  
 AAD21976  
 ID AAD21976 standard; DNA; 3806 BP.  
 XX  
 AC AAD21976;  
 XX  
 DT 12-FEB-2002 (first entry)  
 XX  
 DE Human CD28 DNA.  
 XX  
 KW Human; cytotoxic T lymphocyte antigen-4; CTLA-4; CD28; AIDS; therapy;  
 KW Acquired Immune Deficiency Syndrome; neuroprotective; dermatological;  
 KW immune response; organ transplantation; autoimmune disease; allergy; SLE;  
 KW systemic lupus erythematosus; multiple sclerosis; tumour vaccination;  
 KW immunodeficiency disease; Digeorge Syndrome; cancer; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 223..885  
 FT /\*tag= a  
 FT /product= "Human CD28 protein"  
 XX  
 PN WO200179300-A1.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 12-APR-2001; 2001WO-US012275.  
 XX  
 PR 12-APR-2000; 2000US-0196851P.  
 XX  
 PA (GEMY ) GENETICS INST INC.  
 XX  
 PI Bluestone JA, Collins M, Whitters M, Griffin M, Kranz D;  
 XX  
 DR WPI: 2002-017603/02.  
 DR P-PSDB; AAE13721.  
 XX  
 PT New construct for downmodulating immune response in a subject, has  
 PT exposed surface attached with antigen-binding portion of antibody that  
 PT binds to cytotoxic T lymphocyte antigen-4 and major histocompatibility  
 PT molecule.  
 XX  
 PA (GETH ) GENENTECH INC.

PS Disclosure; Page 93-94; 98pp; English.

XX The invention relates to a construct for downmodulating immune response in a subject. The construct comprises an exposed surface attached with an antigen-binding portion of an antibody that binds to a cytotoxic T lymphocytic antigen (CTLA)-4 or CD28 expressed on T-cell of the subject, and a major histocompatibility complex (MHC) molecule. The construct is useful for treating an individual afflicted with a disease or disorder that would benefit from downregulation of immune response. Downmodulation of the immune response is useful to downmodulate the immune response in situations of tissue, skin and organ transplantation, graft-versus-host disease or in autoimmune diseases such as systemic lupus erythematosus and multiple sclerosis. The construct is useful for inhibiting immune cell activation and prevents production of autoantibodies or cytokines which may be involved in disease process. Inhibition of immune cell activation is useful in the treatment of allergy and allergic reactions e.g., by inhibiting IgE production. The construct is also useful for enhancing desirable immune response in a subject e.g., in situations of tumour vaccination, in viral immunity or in immunodeficiency diseases such as AIDS and DiGeorge Syndrome. Induction/enhancement of immune cell function results in increased tumour destruction in cancer patients. The construct is also useful in treating infectious diseases. The present sequence is human CD28 DNA

XX Sequence 3806 BP; 1079 A; 765 C; 881 G; 1081 T; 0 U; 0 Other;

SQ Query Match 70.4%; Score 19; DB 6; Length 3806;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTCCT 27  
|||||

Db 561 AATTGAAGTTATGTCCT 579  
|||||

RESULT 32

ADA02990

ID ADA02990 standard; DNA; 51365 BP.

XX ADA02990;

XX 06-NOV-2003 (first entry)

DT Human CD28 carcinoma associated gene, SEQ ID NO:1508.

DE Human; carcinoma associated; oncogene; carcinoma; cancer; breast; prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening; gene; ds.

KW Homo sapiens.

XX WO2003057146-A2.

PN 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.

PF 26-DEC-2001; 2001US-00035832.

XX (SAGR-) SAGRES DISCOVERY.

PA Morris DW;

PI WPI; 2003-587068/55.

DR New recombinant nucleic acid encoding carcinoma associated protein, useful for preparing compositions for treating carcinomas.

PT Claim 1; SEQ ID NO 1508; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic acid sequences from mouse and human (ADA01482-ADA03094), and to recombinant carcinoma associated proteins (CAP) encoded by them. The invention also encompasses expression vectors and host cells comprising a CA nucleic acid, a polypeptide (especially an antibody) that specifically binds to the protein, and a biochip comprising CA nucleic acid or fragments thereof. The sequences of the invention were identified using oncogenic retroviruses, which insert into the genome of the host organism at random. Many of these do not carry transduced host oncogenes or pathogenic trans-acting viral genes, meaning that cancer incidence is a direct consequence of the effects of proviral integration into host protooncogenes. The CA nucleic acid sequences can be used to diagnose carcinoma (especially breast cancer, prostate cancer, lymphoma or leukaemia) or a propensity to carcinoma by determination of the sequence of a CA gene, or by determination of CA gene expression in particular tissues. CA nucleic acids, proteins and antibodies are also useful as therapeutic agents and in screening and evaluating drug candidates. The present sequence represents a specifically claimed human CA nucleic acid sequence of the invention. Note: The complete sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 51365 BP; 15176 A; 10059 C; 10474 G; 15656 T; 0 U; 0 Other;

SQ Query Match 70.4%; Score 19; DB 9; Length 51365;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTCCT 27  
|||||

Db 30447 AATTGAAGTTATGTCCT 30465  
|||||

RESULT 33

ADB72728

ID ADB72728 standard; DNA; 51365 BP.

XX ADB72728;

XX 04-DEC-2003 (first entry)

DT Human CD28 gene.

DE human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas; cancer; neoplasm; adenocarcinoma; sarcoma; gene.

KW Homo sapiens.

XX WO2003008583-A2.

PN 30-JAN-2003.

PD 26-DEC-2001; 2001WO-US051291.

PF 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

PR 08-NOV-2001; 2001US-00052482.

PR 30-NOV-2001; 2001US-00997722.

PR 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

PA Morris DW, Engelhard EK;

PI WPI; 2003-239337/23.

DR New recombinant nucleic acid, useful for treating carcinomas, lymphomas, cancers, neoplasm, adenocarcinoma, or sarcomas.

PT Claim 1; SEQ ID NO 556; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a nucleotide sequence selected from any of the 660 sequences fully defined in the specification. A polynucleotide of the invention has cytostatic activity, and may have a use in gene therapy, or in a vaccine. The

```
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 51365 BP; 15176 A; 10059 C; 10474 G; 15656 T; 0 U; 0 Other;
Query Match 70.4%; Score 19; DB 10; Length 51365;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 AATTGAAGTTATGTCCT 27
DB 30447 AATTGAAGTTATGTCCT 30465

RESULT 34
ADC85470
ID ADC85470 standard; DNA; 51365 BP.
XX
AC ADC85470;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human Cd28 genomic sequence.
XX
XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX
OS Homo sapiens.
XX
XX WO2003045230-A2.
XX
PD 05-JUN-2003.
XX
PF 02-DEC-2002; 2002WO-US038582.
XX
PR 30-NOV-2001; 2001US-00997722.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
XX WPI; 2003-513603/48.
XX
XX New recombinant nucleic acid comprising a nucleotide sequence of any of
PT the carcinoma-associated (CA) genes, useful for screening for drug
PT candidates for diagnosing or treating carcinomas.
XX
PS Claim 1; SEQ ID NO 256; 983pp; English.
XX
CC The invention relates to a recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the fully defined carcinoma-
CC associated (CA) genes from the 50 tables given in the specification. The
CC CA proteins are secreted, transmembrane or intracellular proteins. The
CC recombinant nucleic acids are useful for screening for drug candidates
CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC ADC85514 represent CA genes of the invention.
XX
SQ Sequence 51365 BP; 15176 A; 10059 C; 10474 G; 15656 T; 0 U; 0 Other;
Query Match 70.4%; Score 19; DB 10; Length 51365;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 AATTGAAGTTATGTCCT 27
DB 30447 AATTGAAGTTATGTCCT 30465

RESULT 35
ADM74585
ID ADM74585 standard; DNA; 51365 BP.
XX
```

```
AC ADM74585;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human carcinoma associated (CA) nucleic acid #127.
XX
XX Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
XX cytostatic.
XX
OS Homo sapiens.
XX
PN US2004072154-A1.
XX
PD 15-APR-2004.
XX
PF 30-NOV-2001; 2001US-00997722.
XX
PR 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
XX
XX (MORR/) MORRIS D W.
XX (ENG/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
XX WPI; 2004-328562/30.
XX
XX New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
PS Claim 1; SEQ ID NO 256; 29pp; English.
XX
CC The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinoma comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 51365 BP; 15176 A; 10059 C; 10474 G; 15656 T; 0 U; 0 Other;
Query Match 70.4%; Score 19; DB 12; Length 51365;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 9 AATTGAAGTTATGATCCT 27  
 |||||  
 Db 30447 AATTGAAGTTATGATCCT 30465

RESULT 36  
 AAA69169  
 ID AAA69169 standard; DNA; 3692 BP.  
 XX AC AAA69169;  
 XX DT 27-OCT-2000 (first entry)  
 XX Bacteriophage Dp-1 nucleotide sequence dPIORF001.  
 XX Bacteriophage; antimicrobial; genome; identification; antibacterial;  
 KW bacterial growth inhibition; bacterial infection; da.  
 XX Bacteriophage Dp-1.  
 OS WO200032825-A2.  
 XX PN 08-JUN-2000.  
 XX PD 03-DEC-1999; 99WO-IB002040.  
 XX PF 03-DEC-1998; 98US-0110992P.  
 XX PR 03-JUN-1999; 99US-00326144.  
 XX PR 28-SEP-1999; 99US-00407804.  
 XX PR 30-SEP-1999; 99US-0157218P.  
 XX PR 01-DEC-1999; 99US-0168777P.  
 XX PR 02-DEC-1999; 99US-00454252.  
 XX (PHAG-) PHAGETECH INC.  
 PA Pelletier J, Gros P, Dubow M;  
 PI WPI; 2000-412361/35.  
 XX P-PSDB; AAB16682.  
 XX Identifying a bacteriophage coding region for treating bacterial  
 PT infections comprises identifying a nucleic acid encoding a product that  
 PT inhibits bacteria when a bacteriophage infects a bacterium.  
 XX Example 17; Page 363-364; 456pp; English.  
 XX The present invention describes a method for identifying a bacteriophage  
 CC coding region encoding a product active on an essential bacterial target.  
 CC The method comprises identifying a nucleic acid sequence encoding a gene  
 CC product that provides a bacteria-inhibiting function when an  
 CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
 CC compound active on a target of a bacteriophage inhibitor protein in a  
 CC bacteria is used to treat or prevent a bacterial infection in an animal.  
 CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage  
 CC nucleotide and protein sequences which are used in the exemplification of  
 CC the present invention  
 XX SQ Sequence 3692 BP; 1151 A; 725 C; 842 G; 974 T; 0 U; 0 Other;  
 Query Match 67.4%; Score 18.2; DB 3; Length 3692;  
 Best Local Similarity 87.0%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GCCGCAATTGAAGTTATGATCC 26  
 |||||  
 Db 1582 GCCGCAACTGAAGTCATGATGC 1504

RESULT 37  
 AAA69168  
 ID AAA69168 standard; DNA; 56506 BP.  
 XX AC AAA69168;  
 XX DT 05-OCT-2001 (first entry)  
 XX Human foetal cDNA, SEQ ID NO: 742.  
 XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;  
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
 KW gene therapy; antisense therapy; cancer; immune disorder;  
 KW growth disorder; osteoporosis; thrombolytic disorder;  
 KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.

AC AAA69168;  
 XX 27-OCT-2000 (first entry)  
 XX Bacteriophage Dp-1 complete genome sequence.  
 DE Bacteriophage; antimicrobial; genome; identification; antibacterial;  
 KW bacterial growth inhibition; bacterial infection; da.  
 XX Bacteriophage Dp-1.  
 OS WO200032825-A2.  
 XX PN 08-JUN-2000.  
 XX PD 03-DEC-1999; 99WO-IB002040.  
 XX PF 03-DEC-1998; 98US-0110992P.  
 XX PR 03-JUN-1999; 99US-00326144.  
 XX PR 28-SEP-1999; 99US-00407804.  
 XX PR 30-SEP-1999; 99US-0157218P.  
 XX PR 01-DEC-1999; 99US-0168777P.  
 XX PR 02-DEC-1999; 99US-00454252.  
 XX (PHAG-) PHAGETECH INC.  
 PA Pelletier J, Gros P, Dubow M;  
 PI WPI; 2000-412361/35.  
 XX Identifying a bacteriophage coding region for treating bacterial  
 PT infections comprises identifying a nucleic acid encoding a product that  
 PT inhibits bacteria when a bacteriophage infects a bacterium.  
 XX Example 17; Page 348-358; 456pp; English.  
 XX The present invention describes a method for identifying a bacteriophage  
 CC coding region encoding a product active on an essential bacterial target.  
 CC The method comprises identifying a nucleic acid sequence encoding a gene  
 CC product that provides a bacteria-inhibiting function when an  
 CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
 CC compound active on a target of a bacteriophage inhibitor protein in a  
 CC bacteria is used to treat or prevent a bacterial infection in an animal.  
 CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage  
 CC nucleotide and protein sequences which are used in the exemplification of  
 CC the present invention  
 XX SQ Sequence 56506 BP; 18170 A; 10773 C; 12019 G; 15544 T; 0 U; 0 Other;  
 Query Match 67.4%; Score 18.2; DB 3; Length 56506;  
 Best Local Similarity 87.0%; Pred. No. 2.1e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GCCGCAATTGAAGTTATGATCC 26  
 |||||  
 Db 38279 GCCGCAACTGAAGTCATGATGC 38301

RESULT 38  
 AAH94213  
 ID AAH94213 standard; cDNA; 392 BP.  
 XX AC AAH94213;  
 XX 05-OCT-2001 (first entry)  
 XX Human foetal cDNA, SEQ ID NO: 742.  
 XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;  
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
 KW gene therapy; antisense therapy; cancer; immune disorder;  
 KW growth disorder; osteoporosis; thrombolytic disorder;  
 KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.

```
XX OS Homo sapiens.
XX PN WO200155339-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US0002723.
XX PR 25-JAN-2000; 2000US-00491404.
XX PR 15-SEP-2000; 2000US-00663870.
XX PR 06-NOV-2000; 2000US-00707351.
XX PA (HYSE-) HYSEQ INC.
XX PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
PI Liu C, Asundi V, Zhou P, Werhman T;
XX WPI; 2001-465571/50.
XX DR P-PSDB; AAM06538.
XX PT Novel fetal proteins useful for the treatment and diagnosis of diseases
XX PT associated with dysfunction of the protein e.g. cancers, immune
XX PT disorders, growth disorders, thrombolytic disorders, nervous system
XX PT disorders and inflammation.
XX PS Claim 1; Page 458; 715pp; English.
XX CC The invention relates to novel foetal polypeptides encoded by
XX CC polynucleotides comprising one of 477 sequences fully defined in the
XX CC specification. The foetal polynucleotides and polypeptides are useful in
XX CC the treatment and diagnosis of diseases such as cancers, immune
XX CC disorders, growth disorders (e.g. osteoporosis), thrombolytic disorders,
XX CC nervous system disorders and inflammation. The present sequence was
XX CC assembled using an expressed sequence tag (EST) found to be expressed in
XX CC human foetal tissue cDNA libraries as the seed
XX SQ Sequence 392 BP; 104 A; 93 C; 76 G; 119 T; 0 U; 0 Other;
Query Match 66.7%; Score 18; DB 5; Length 392;
Best Local Similarity 80.8%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 CGGCCGAATTGAAGTTATGATTCCT 27
DB 113 CTGCTGCAATTCAAGTTATTCCT 138
RESULT 39
ADJ41623
ID ADJ41623 standard; cDNA; 2000 BP.
XX AC ADJ41623;
XX DT 06-MAY-2004 (first entry)
XX DE Plant cDNA #2623.
XX KW plant; gene; ss; transcription; plant genome augmentation; cereal;
XX KW soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
XX KW maize; barley; sorghum; rice; wheat; cotton plant; insecticide resistance;
XX KW stress tolerance; salt tolerance; cold tolerance; drought tolerance;
XX KW plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
XX KW antifungal.
XX OS Eukaryota.
XX PN US2004016025-A1.
XX PD 22-JAN-2004.
XX PF 26-SEP-2002; 2002US-00260238.
XX
```

```
PR 26-SEP-2001; 2001US-0325277P.
PR 26-SEP-2001; 2001US-0325448P.
PR 04-APR-2002; 2002US-0370620P.
XX PA (BUDW/) BUDWORTH P.
XX PA (MOUG/) MOUGHAMER T.
XX PA (BRIG/) BRIGGS S P.
XX PA (COOP/) COOPER B.
XX PA (GLAZ/) GLAZEBROOK J.
XX PA (GOFF/) GOFF S A.
XX PA (KATA/) KATAGIRI F.
XX PA (KREP/) KRSPS J.
XX PA (PROV/) PROVART N.
XX PA (RICK/) RICKS D.
XX PA (ZHUT/) ZHU T.
XX PI Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI Goff SA, Katagiri F, Krep J, Provart N, Ricke D, Zhu T;
XX WPI; 2004-190374/18.
XX PT New rice promoter, useful for manipulating crop plants to alter or
XX PT improve phenotypic characteristics, e.g. produce large quantities of oil
XX PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
XX PT or high nutritional value.
XX PS Claim 1; SEQ ID NO 2623; 230pp; English.
XX CC The invention relates to plant nucleotide sequences that direct seed-,
XX CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
XX CC or constitutive transcription of an operatively linked nucleic acid
XX CC segment. The invention also relates to a method for augmenting a plant
XX CC genome and a method of identifying a gene, where its expression is
XX CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
XX CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
XX CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
XX CC sorghum, rice or wheat. The polynucleotides and the polypeptides they
XX CC encode are useful for manipulating crop plants to alter or improve
XX CC phenotypic characteristics, to produce large quantities of oil or
XX CC proteins, to incur resistance to insecticides, viruses or fungi, and to
XX CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
XX CC have a high nutritional value with reduced apical dominance or dwarfism,
XX CC early flowering or altered metabolic pathways. This sequence represents a
XX CC plant nucleic acid of the invention. Note: The sequence data for this
XX CC patent did not form part of the printed specification but was obtained in
XX CC electronic format directly from USPTO at segdata.uspto.gov/sequence.html.
XX SQ Sequence 2000 BP; 491 A; 502 C; 447 G; 559 T; 0 U; 1 Other;
Query Match 66.7%; Score 18; DB 12; Length 2000;
Best Local Similarity 80.8%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GGGCGCGCAATTGAGTTATGATTC 26
DB 1484 GCAGCGCAATCGAAGGTGTGATCC 1509
RESULT 40
ABL24354/c
ID ABL24354 standard; DNA; 4093 BP.
XX AC ABL24354;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 24535.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX
```

PN WO200171042-A2.  
XX  
PD  
XX  
XX 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
DR  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 24535; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 4093 BP; 1185 A; 920 C; 888 G; 1100 T; 0 U; 0 Other;  
Query Match 66.7%; Score 18; DB 4; Length 4093;  
Best Local Similarity 80.8%; Pred. No. 1.7e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
OY 2 CGGCCGCAATTGAAGTTATGATCCT 27  
|||||  
DB 2476 CGGCTTCATTGAAGTCAGATATCCT 2451  
|||||

Search completed: January 7, 2005, 11:40:47  
Job time : 209.579 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 48.7895 Seconds  
(without alignments)  
393.349 Million cell updates/sec

Title: US-09-786-502A-7

Perfect score: 27  
Sequence: 1 ggggcccgaattgaagttatgcctc 27

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	70.4	660	3	US-08-812-946A-1
2	19	70.4	1514	4	US-09-023-655-1316
3	18	66.7	1664976	4	US-08-916-421B-1
4	18	66.7	1664976	4	US-09-692-570-1
5	17.8	65.9	2040	4	US-09-107-532A-2735
6	17.6	65.2	40	4	US-09-702-498A-3
7	17.6	65.2	40	4	US-09-702-498A-4
8	17.6	65.2	16011	4	US-09-600-319-3
9	17.4	64.4	6176	4	US-09-976-594-224
10	17.4	64.4	6290	4	US-09-919-039-106
11	17	63.0	130	4	US-09-536-977-13
12	17	63.0	131	1	US-08-324-243-17
13	17	63.0	131	1	US-08-532-390-17
14	17	63.0	131	3	US-08-717-294-17
15	17	63.0	131	5	PCT-US95-11511-17
16	17	63.0	374	4	US-09-536-977-43
17	17	63.0	603	4	US-09-134-000C-196
18	17	63.0	1277	4	US-09-536-977-45
19	17	63.0	1277	4	US-09-536-977-47
20	17	63.0	1277	4	US-09-536-977-49
21	17	63.0	1277	4	US-08-536-977-51
22	17	63.0	1323	4	US-09-475-515-35
23	17	63.0	1387	4	US-09-475-515-34
24	17	63.0	1431	4	US-09-536-977-73
25	17	63.0	1453	4	US-09-475-515-33
26	17	63.0	1632	1	US-08-324-243-34
27	17	63.0	1632	1	US-08-532-390-34

c 28	17	63.0	1632	3	US-08-717-294-34	Sequence 34, Appl
c 29	17	63.0	1632	5	PCT-US95-11511-34	Sequence 34, Appl
c 30	17	63.0	1836	4	US-09-475-515-41	Sequence 41, Appl
c 31	17	63.0	1836	4	US-09-475-515-44	Sequence 44, Appl
c 32	17	63.0	1836	4	US-09-475-515-47	Sequence 47, Appl
c 33	17	63.0	1918	4	US-09-536-977-67	Sequence 67, Appl
c 34	17	63.0	1944	4	US-09-475-515-37	Sequence 37, Appl
c 35	17	63.0	1944	4	US-09-475-515-38	Sequence 38, Appl
c 36	17	63.0	1944	4	US-09-475-515-40	Sequence 40, Appl
c 37	17	63.0	1944	4	US-09-475-515-43	Sequence 43, Appl
c 38	17	63.0	1944	4	US-09-475-515-46	Sequence 46, Appl
c 39	17	63.0	2025	4	US-09-475-515-36	Sequence 36, Appl
c 40	17	63.0	2025	4	US-09-475-515-39	Sequence 39, Appl
c 41	17	63.0	2025	4	US-09-475-515-42	Sequence 42, Appl
c 42	17	63.0	2025	4	US-09-475-515-45	Sequence 45, Appl
c 43	17	63.0	2071	4	US-09-536-977-69	Sequence 69, Appl
c 44	17	63.0	2079	4	US-09-248-796A-5314	Sequence 5314, Ap
c 45	17	63.0	2298	4	US-09-476-242-22	Sequence 22, Appl

## ALIGNMENTS

## RESULT 1

US-08-812-946A-1

; Sequence 1, Application US/08812946A

; Patent No. 6221637

; GENERAL INFORMATION:

; APPLICANT: Tsuneaki HIDA et al.

; TITLE OF INVENTION: XANTHENE DERIVATIVES, THEIR PRODUCTION AND

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack

; STREET: 805 Fifteenth Street, N.W., #700

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/812,946A

; FILING DATE: March 4, 1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Warren M. Cheek, Jr.

; REGISTRATION NUMBER: 33,367

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-8850

; TELEFAX:

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 660 bases

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-812-946A-1

Query Match 70.4%; Score 19; DB 3; Length 660;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAGTTATGTCCT 27

Db 339 AATTGAAGTTATGTCCT 357  
|||||  
RESULT 2  
US-09-023-655-1316  
; Sequence 1316, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Sellhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1316:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1514 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: g338444  
; US-09-023-655-1316  
  
Query Match 70.4%; Score 19; DB 4; Length 1514;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 AATTGAAGTTATGTCCT 27  
|||||  
Db 438 AATTGAAGTTATGTCCT 456  
|||||  
RESULT 3  
US-08-916-421B-1  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:  
; APPLICANT: Bult et al.  
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco  
; Patent No. 6503729  
; TITLE OF INVENTION: jannaschii  
; FILE REFERENCE: PB275  
; CURRENT APPLICATION NUMBER: US/08/916,421B

```

; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc.feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc.feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc.feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc.feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc.feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc.feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

**Qy**            1 GCGGCCGAATTGAGCTTATGTATCC 26  
               |||      |||      |      |||||      |||  
**Db**          609314 GCAGCAGCTATTGATGATGTATCC 609339

US-09-692-570-1

```

US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genom
; Patent No. 6797466
; TITLE OF INVENTION: jannaschi
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/02
; PRIOR FILING DATE: 1995-08-22
; PRIOR APPLICATION NUMBER: US 08/91
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschi
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t,
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t,
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t,
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t,
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (94812)..(94812)
; OTHER INFORMATION: n equals a, t,

```

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (98343)..(98343)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (103998)..(103998)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (191995)..(191995)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (231380)..(231380)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (234187)..(234187)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (234220)..(234220)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:

NAME/KEY: misc feature  
LOCATION: (319226)..(319226)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (559167)..(559167)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (559241)..(559241)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (600992)..(600992)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (622708)..(622708)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (657081)..(657081)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (657203)..(657203)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (674435)..(674435)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (682442)..(682442)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (713652)..(713652)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (741684)..(741684)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (779455)..(779455)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (779676)..(779676)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature

```
; LOCATION: (1130881)...(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)...(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)...(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)...(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match      66.7%; Score 18; DB 4; Length 1664976;
Best Local Similarity 80.8%; Pred. No. 78;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GCGGCGCAATTCGAAGTTATGTATCC 26
Db      609314 GCACGAGCTATGATGATGTATCC 609339

RESULT 5
US-09-107-532A-2735
; Sequence 2735, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2735:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2040 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (B) LOCATION 1...2040
; SEQUENCE DESCRIPTION: SEQ ID NO: 2735:
US-09-107-532A-2735

Query Match      65.9%; Score 17.8; DB 4; Length 2040;
Best Local Similarity 90.5%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      7 GCAATTGAAGTTATGTATCCT 27
Db      1060 GCAATCGAAGTTATTATCCT 1080

RESULT 6
US-09-702-498A-3
; Sequence 3, Application US/09702498A
; Patent No. 6746860
; GENERAL INFORMATION:
; APPLICANT: Tokusumi, Tsuyoshi
; APPLICANT: Iida, Akihiro
; APPLICANT: Hasegawa, Mamoru
; APPLICANT: Nagai, Yoshiyuki
; TITLE OF INVENTION: Paramyxovirus vectors used for transfer
; FILE REFERENCE: 50026/025001
; CURRENT APPLICATION NUMBER: US/09/702,498A
; CURRENT FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: JP 2000-152726
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially synthesized sequence derived from the
; OTHER INFORMATION: Sendai virus
US-09-702-498A-3

Query Match      65.2%; Score 17.6; DB 4; Length 40;
Best Local Similarity 83.3%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GCGGCGCAATTCGAAGTTATGTAT 24
Db      15 GCGGCGCAATGCGAGATATCTAT 38

RESULT 7
US-09-702-498A-4/c
; Sequence 4, Application US/09702498A
; Patent No. 6746860
; GENERAL INFORMATION:
; APPLICANT: Tokusumi, Tsuyoshi
; APPLICANT: Iida, Akihiro
; APPLICANT: Hasegawa, Mamoru
; APPLICANT: Nagai, Yoshiyuki
; TITLE OF INVENTION: Paramyxovirus vectors used for transfer
; FILE REFERENCE: 50026/025001
; CURRENT APPLICATION NUMBER: US/09/702,498A
; CURRENT FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: JP 2000-152726
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially synthesized sequence derived from the
```

```
; OTHER INFORMATION: Sendai virus
US-09-702-498A-4

Query Match      65.2%; Score 17.6; DB 4; Length 40;
Best Local Similarity 83.3%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGCGCCGAATTGAAGTTATGTAT 24
Db 26 GCGCGCCGAATGGCAGATATCTAT 3

RESULT 8
US-09-600-319-3/c
; Sequence 3, Application US/09600319
; Patent No. 6780610
; GENERAL INFORMATION:
; APPLICANT: Owens, Gary
; APPLICANT: Madsen, Cort
; TITLE OF INVENTION: Identification of a Smooth Muscle Cell (SMC) Specific Smooth Muscle
; FILE OF INVENTION: Myosin Heavy Chain (SM-MHC) Promoter/Enhancer
; FILE REFERENCE: 00241-03
; CURRENT APPLICATION NUMBER: US/09/600,319
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: PCT/US99/01038
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: 60/071,300
; PRIOR FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 16011
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-600-319-3

Query Match      65.2%; Score 17.6; DB 4; Length 16011;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GCGCGCAATTGAAGTTATGTATCCT 27
Db 6333 GCTGGCAATTGAAGTTATGTACCAT 6310

RESULT 9
US-09-976-594-224
; Sequence 224, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 224
; LENGTH: 6176
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 282397.77
US-09-976-594-224

Query Match      64.4%; Score 17.4; DB 4; Length 6176;
Best Local Similarity 77.8%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 GCGCGCCGAATTGAAGTTATGTATCCT 27
Db 4060 GCGCCAGCAGTTGAAGCTATACATCCT 4086

RESULT 10
US-09-919-039-106/c
; Sequence 106, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaefer, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 106
; LENGTH: 6290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6727066 282397.85c
US-09-919-039-106

Query Match      64.4%; Score 17.4; DB 4; Length 6290;
Best Local Similarity 77.8%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCGCGCCGAATTGAAGTTATGTATCCT 27
Db 2117 GCGCCAGCAGTTGAAGCTATACATCCT 2091

RESULT 11
US-09-536-977-13/c
; Sequence 13, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; FILE REFERENCE: BX08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 130
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(130)
US-09-536-977-13

Query Match      63.0%; Score 17; DB 4; Length 130;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCCGAATTGAAGTTATGTATC 25
Db 52 GCGCGCCAGTTGAAGCTGTGCATC 28
```

RESULT 12  
US-08-324-243-17/c  
; Sequence 17, Application US/08324243  
; Patent No. 5786464  
; GENERAL INFORMATION:  
; APPLICANT: SEED, BRIAN  
; TITLE OF INVENTION: OVEREXPRESSION OF MAMMALIAN AND VIRAL  
; TITLE OF INVENTION: PROTEINS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/324,243  
; FILING DATE: 19-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CLARK, PAUL T  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/226001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-324-243-17

Query Match 63.0%; Score 17; DB 1; Length 131;  
Best Local Similarity 80.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25  
DB 131 GCGCGCGCAGTTGAAGCTGTGCATC 107

RESULT 13  
US-08-532-390-17/c  
; Sequence 17, Application US/08532390  
; Patent No. 5795737  
; GENERAL INFORMATION:  
; APPLICANT: SEED, BRIAN  
; APPLICANT: HAAS, JURGEN  
; TITLE OF INVENTION: High Level Expression of Proteins  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/532,390

; FILING DATE: 22-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/324,243  
; FILING DATE: 19-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LECH, KAREN F.  
; REGISTRATION NUMBER: 35,238  
; REFERENCE/DOCKET NUMBER: 00786/294001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-532-390-17

Query Match 63.0%; Score 17; DB 1; Length 131;  
Best Local Similarity 80.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25  
DB 131 GCGCGCGCAGTTGAAGCTGTGCATC 107

RESULT 14  
US-08-717-294-17/c  
; Sequence 17, Application US/08717294  
; Patent No. 6114148  
; GENERAL INFORMATION:  
; APPLICANT: SEED, BRIAN  
; APPLICANT: HAAS, JURGEN  
; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF  
; TITLE OF INVENTION: PROTEINS  
; NUMBER OF SEQUENCES: 110  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Clark & Elbing LLP  
; STREET: 176 Federal Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/717,294  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elbing, Karen L  
; REGISTRATION NUMBER: 35,238  
; REFERENCE/DOCKET NUMBER: 00786/345001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-428-0200  
; TELEFAX: 617-428-7045  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Other

US-08-717-294-17

```
Query Match      63.0%; Score 17; DB 3; Length 131;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 5; Indels
```

**Qy**            1 GCGGCCGCAATTGAAGTTATGTATC 25  
               |||      |      |      |      |      |  
**Dβ**            131 GCGCCGCAGTTGAAGCTGTGCATC 107

RESULT 15

```

PCT-US95-11511-17/c
; Sequence 17, Application PC/TUS9511511
; GENERAL INFORMATION:
; APPLICANT: SEED, BRIAN
; TITLE OF INVENTION: OVEREXPRESSION OF MAMMALIAN AND VIRAL
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11511
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, PAUL T
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/226001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US95-11511-17

```

```
Query Match      63.0%; Score 17; DB 5; Length 131;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 20: Conservative 0; Mismatches 5; Indels
```

QY 1 GCGGCCGCAATTGAAGTTATGTATC 25  
Dp 131 GCGGCCGCAAGTTGAAGCTGTGCATC 107

RESULT 16

```

RESOLUTION 18
US-09-536-977-43/c
; Sequence 43, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; TITLE OF INVENTION: BX08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558

```

```

; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 374
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(372)
US-09-536-977-43

```

Query Match 63.0%; Score 17; DB 4; Length 374;  
Best Local Similarity 80.0%; Pred. No. 44;  
Matches 20; Conservative 0; Mismatches 5; Indels

Qy 1 GCGCCGCAATTGAAGTTATGTATC 25  
|||  
Db 138 GCGCCGCGAGTTGAAGCTGTGCATC 114

RESULT 17

```

US-09-134-000C-196
; Sequence 196, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 196
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-196

```

Query Match 63.0%; Score 17; DB 4; Length 603;  
Best Local Similarity 80.0%; Pred.No. 49;  
Matches 20; Conservative 0; Mismatches 5; Indels

QY 1 GCGGCCGCAATTGAAGTTATGTATC 25  
|||||  
Dp 538 GCGGCTGAATTTGAAGTATTATCC 562

RESUL, T 18

```

RESOL 18
US-09-536-977-45/c
; Sequence 45, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOWSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; TITLE OF INVENTION: BX08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45

```



```
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1275)
US-09-536-977-45

Query Match      63.0%; Score 17; DB 4; Length 1277;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25
Db 1041 GCGCGCGCAGTTGAAGCTGTGCATC 1017

RESULT 19
US-09-536-977-47/c
; Sequence 47, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1275)
US-09-536-977-47

Query Match      63.0%; Score 17; DB 4; Length 1277;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25
Db 1041 GCGCGCGCAGTTGAAGCTGTGCATC 1017

RESULT 20
US-09-536-977-49/c
; Sequence 49, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1275)
US-09-536-977-49

Query Match      63.0%; Score 17; DB 4; Length 1277;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25
Db 1041 GCGCGCGCAGTTGAAGCTGTGCATC 1017

RESULT 21
US-09-536-977-51/c
; Sequence 51, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1275)
US-09-536-977-51

Query Match      63.0%; Score 17; DB 4; Length 1277;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25
Db 1041 GCGCGCGCAGTTGAAGCTGTGCATC 1017

RESULT 22
US-09-475-515-35/c
; Sequence 35, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1323
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp120.modSF162.delV1V2
US-09-475-515-35

Query Match      63.0%; Score 17; DB 4; Length 1323;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCGCAATTGAAGTTATGTATC 25
Db 936 GCGGCGCGCAGTTGAAGCTGTGCATC 912

RESULT 23
US-09-475-515-34/c
; Sequence 34, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp120.modSF162.delV2
US-09-475-515-34

Query Match      63.0%; Score 17; DB 4; Length 1387;
Best Local Similarity 80.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCGCAATTGAAGTTATGTATC 25
Db 1044 GCGGCGCGCAGTTGAAGCTGTGCATC 1020

RESULT 24
US-09-536-977-73/c
; Sequence 73, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; TITLE OF INVENTION: BX08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 73
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1428)
US-09-536-977-73

Query Match      63.0%; Score 17; DB 4; Length 1431;
Best Local Similarity 80.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCGCAATTGAAGTTATGTATC 25
Db 1041 GCGGCGCGCAGTTGAAGCTGTGCATC 1017

RESULT 25
US-09-475-515-33/c
; Sequence 33, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33
; LENGTH: 1453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: gp120.modSF162
US-09-475-515-33

Query Match      63.0%; Score 17; DB 4; Length 1453;
Best Local Similarity 80.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCGCAATTGAAGTTATGTATC 25
Db 1125 GCGGCGCGCAGTTGAAGCTGTGCATC 1101

RESULT 26
US-08-324-243-34/c
; Sequence 34, Application US/08324243
; Patent No. 5786464
; GENERAL INFORMATION:
; APPLICANT: SEED, BRIAN
; TITLE OF INVENTION: OVEREXPRESSION OF MAMMALIAN AND VIRAL
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/324,243  
FILING DATE: 19-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: CLARK, PAUL T  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/226001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1632 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-324-243-34

Query Match 63.0%; Score 17; DB 1; Length 1632;  
Best Local Similarity 80.0%; Pred. No. 61;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25  
Db 1230 GCGCGCGCAGTTGAAGCTGTGCATC 1206

RESULT 27  
US-08-532-390-34/c  
Sequence 34, Application US/08532390  
Patent No. 5795737  
GENERAL INFORMATION:  
APPLICANT: SEED, BRIAN  
APPLICANT: HAAS, JURGEN  
TITLE OF INVENTION: High Level Expression of Proteins  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,390  
FILING DATE: 22-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/324,243  
FILING DATE: 19-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: LECH, KAREN F.  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/294001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1632 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-532-390-34

Query Match 63.0%; Score 17; DB 1; Length 1632;  
Best Local Similarity 80.0%; Pred. No. 61;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25  
Db 1230 GCGCGCGCAGTTGAAGCTGTGCATC 1206

RESULT 28  
US-08-717-294-34/c  
Sequence 34, Application US/08717294  
Patent No. 6114148  
GENERAL INFORMATION:  
APPLICANT: SEED, BRIAN  
APPLICANT: HAAS, JURGEN  
TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF  
TITLE OF INVENTION: PROTEINS  
NUMBER OF SEQUENCES: 110  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Clark & Elbing LLP  
STREET: 176 Federal Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/717,294  
FILING DATE: 20-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Elbing, Karen L  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/345001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-428-0200  
TELEFAX: 617-428-7045  
TELEX:  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1632 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-717-294-34

Query Match 63.0%; Score 17; DB 3; Length 1632;  
Best Local Similarity 80.0%; Pred. No. 61;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25  
Db 1230 GCGCGCGCAGTTGAAGCTGTGCATC 1206

RESULT 29  
PCT-US95-11511-34/c  
Sequence 34, Application PC/TUS9511511  
GENERAL INFORMATION:  
APPLICANT: SEED, BRIAN  
TITLE OF INVENTION: OVEREXPRESSION OF MAMMALIAN AND VIRAL  
TITLE OF INVENTION: PROTEINS

```
;
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11511
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, PAUL T
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/226001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1632 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US95-11511-34

Query Match 63.0%; Score 17; DB 5; Length 1632;
Best Local Similarity 80.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
Db 1230 GCGGCGCAGTTGAAGCTGTGCATC 1206

RESULT 30
US-09-475-515-41/c
; Sequence 41, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut.modSF162.delV1/V2
US-09-475-515-41

Query Match 63.0%; Score 17; DB 4; Length 1836;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
Db 936 GCGGCGCAGTTGAAGCTGTGCATC 912

RESULT 32
US-09-475-515-47/c
; Sequence 47, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut.modSF162.delV1/V2
US-09-475-515-41

Query Match 63.0%; Score 17; DB 4; Length 1836;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
Db 936 GCGGCGCAGTTGAAGCTGTGCATC 912

RESULT 31
US-09-475-515-44/c
; Sequence 44, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut7.modSF162.delV1/V2
US-09-475-515-44

Query Match 63.0%; Score 17; DB 4; Length 1836;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
Db 936 GCGGCGCAGTTGAAGCTGTGCATC 912

RESULT 33
US-09-475-515-45/c
; Sequence 45, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut.modSF162.delV1/V2
US-09-475-515-45
```

```
; OTHER INFORMATION: gp140.mut8.modSF162.delv1/v2
US-09-475-515-47

Query Match      63.0%; Score 17; DB 4; Length 1836;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
   |||||||
Db 936 GCGGCGCAGTTGAAGCTGTGCATC 912

RESULT 33
US-09-536-977-67/c
; Sequence 67, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; TITLE OF INVENTION: BK08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
; LENGTH: 1918
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1905)
US-09-536-977-67

Query Match      63.0%; Score 17; DB 4; Length 1918;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
   |||||||
Db 1041 GCGGCGCAGTTGAAGCTGTGCATC 1017

RESULT 34
US-09-475-515-37/c
; Sequence 37, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1944
; TYPE: DNA

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.modSF162.delv1/v2
US-09-475-515-37

Query Match      63.0%; Score 17; DB 4; Length 1944;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
   |||||||
Db 1044 GCGGCGCAGTTGAAGCTGTGCATC 1020

RESULT 35
US-09-475-515-38/c
; Sequence 38, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.modSF162.delv1/v2
US-09-475-515-38

Query Match      63.0%; Score 17; DB 4; Length 1944;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATC 25
   |||||||
Db 1044 GCGGCGCAGTTGAAGCTGTGCATC 1020

RESULT 36
US-09-475-515-40/c
; Sequence 40, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
```

```
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut.modsf162.delv2
US-09-475-515-40

Query Match      63.0%; Score 17; DB 4; Length 1944;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCCGCAATTGAAGTTATGTATC 25
Db 1044 GCCGCCGAGTTGAAGCTGTGCATC 1020

RESULT 37
US-09-475-515-43/c
; Sequence 43, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut7.modsf162.delv2
US-09-475-515-43

Query Match      63.0%; Score 17; DB 4; Length 1944;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCCGCAATTGAAGTTATGTATC 25
Db 1044 GCCGCCGAGTTGAAGCTGTGCATC 1020

RESULT 38
US-09-475-515-46/c
; Sequence 46, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut8.modsf162.delv2
US-09-475-515-46

Query Match      63.0%; Score 17; DB 4; Length 1944;
Best Local Similarity 80.0%; Pred. No. 64;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCCGCAATTGAAGTTATGTATC 25
Db 1125 GCCGCCGAGTTGAAGCTGTGCATC 1101

RESULT 39
US-09-475-515-36/c
; Sequence 36, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2025
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.modsf162
US-09-475-515-36

Query Match      63.0%; Score 17; DB 4; Length 2025;
Best Local Similarity 80.0%; Pred. No. 64;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCCGCAATTGAAGTTATGTATC 25
Db 1125 GCCGCCGAGTTGAAGCTGTGCATC 1101

RESULT 40
US-09-475-515-39/c
; Sequence 39, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
```

```

; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2025
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp140.mut.modsf162
US-09-475-515-39

Query Match      63.0%; Score 17; DB 4; Length 2025;
Best Local Similarity 80.0%; Pred. No. 64;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GCGCGCGCAATTGAGTTATGTATC 25
Db      1125 GCGCGCGCAGTTGAGCTGTGCATC 1101

Search completed: January 7, 2005, 11:00:35
Job time : 53.7895 secs

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 11:00:44 ; Search time 1147.74 Seconds  
(without alignments)  
134.979 Million cell updates/sec

Title: US-09-786-502A-7  
Perfect score: 27  
Sequence: 1 gggcgcaattgaattatgtatcct 27

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4293498 seqs, 2868903791 residues

Total number of hits satisfying chosen parameters: 8586996

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq.\*
- 20: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq.\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	27	100.0	27	8	US-08-940-544-1
2	19	70.4	328	16	US-10-448-256-6
3	19	70.4	663	11	US-09-997-722-258
4	19	70.4	1514	9	US-09-738-546-1
5	19	70.4	1514	11	US-09-836-544-7
6	19	70.4	1514	15	US-10-207-655-98
7	19	70.4	1514	16	US-10-641-643-1316
8	19	70.4	3803	15	US-10-076-934-1
9	19	70.4	3804	11	US-09-997-722-257
10	19	70.4	3804	15	US-10-143-238-1
11	19	70.4	3806	10	US-09-835-297-3
12	19	70.4	51365	11	US-09-997-722-256

13	18.2	67.4	3693	15	US-10-097-111-11	Sequence 11, Appl
14	18.2	67.4	56506	15	US-10-097-111-10	Sequence 10, Appl
15	18	66.7	1112	17	US-10-437-963-96729	Sequence 96729, A
16	18	66.7	2000	16	US-10-260-238-2623	Sequence 2623, A
17	17.8	65.9	2178	16	US-10-282-122A-21821	Sequence 21821, A
18	17.8	65.9	3624	17	US-10-697-828-23	Sequence 23, Appl
19	17.8	65.9	4121	17	US-10-697-828-19	Sequence 19, Appl
20	17.8	65.9	5084	17	US-10-697-828-16	Sequence 16, Appl
21	17.8	65.9	48445	13	US-10-087-192-25	Sequence 25, Appl
22	17.6	65.2	40	9	US-09-966-277-21	Sequence 21, Appl
23	17.6	65.2	40	9	US-09-966-277-22	Sequence 22, Appl
24	17.6	65.2	40	10	US-09-966-930-21	Sequence 21, Appl
25	17.6	65.2	40	10	US-09-966-930-22	Sequence 22, Appl
26	17.6	65.2	40	15	US-10-316-538-21	Sequence 21, Appl
27	17.6	65.2	40	15	US-10-316-538-22	Sequence 22, Appl
28	17.6	65.2	40	15	US-10-316-535-21	Sequence 21, Appl
29	17.6	65.2	40	15	US-10-316-535-22	Sequence 22, Appl
30	17.6	65.2	40	17	US-10-784-710-3	Sequence 3, Appl
31	17.6	65.2	40	17	US-10-784-710-4	Sequence 4, Appl
32	17.6	65.2	2006	17	US-10-437-963-33246	Sequence 33246, A
33	17.6	65.2	2541	16	US-10-264-049-104	Sequence 104, Appl
34	17.6	65.2	16011	14	US-10-057-726-16	Sequence 16, Appl
35	17.4	64.4	41	10	US-09-769-736-186	Sequence 186, Appl
36	17.4	64.4	154	11	US-09-987-899-5599	Sequence 5599, Ap
37	17.4	64.4	285	11	US-09-987-899-5599	Sequence 5599, Ap
38	17.4	64.4	292	16	US-10-242-535A-1095	Sequence 1095, Ap
39	17.4	64.4	292	16	US-10-085-783A-1095	Sequence 1095, Ap
40	17.4	64.4	514	15	US-10-029-386-1391	Sequence 1391, Ap
41	17.4	64.4	870	16	US-10-282-122A-15342	Sequence 15342, A
42	17.4	64.4	1029	16	US-10-425-114-4668	Sequence 4668, Ap
43	17.4	64.4	1208	13	US-10-027-632-264609	Sequence 264609, Sequence 264609,
44	17.4	64.4	1208	15	US-10-027-632-264609	Sequence 264609,
45	17.4	64.4	1714	18	US-10-425-115-68262	Sequence 68262, A

ALIGNMENTS

RESULT 1

US-08-940-544-1  
; Sequence 1, Application US/08940544B  
; Publication No. US20020018783A1  
; GENERAL INFORMATION:  
; APPLICANT: SADELAIN, MICHEL  
; APPLICANT: CHEUNG, NAI-KONG V.  
; APPLICANT: KHAUSE, ANJA  
; APPLICANT: GUO, HONG-FEN  
; TITLE OF INVENTION: FUSION PROTEINS OF A SINGLE CHAIN ANTIBODY AND CD28 AND  
; FILE REFERENCE: US/08-940-544B  
; CURRENT APPLICATION NUMBER: US/08/940,544B  
; CURRENT FILING DATE: 1997-09-30  
; EARLIER APPLICATION NUMBER: PCT/US97/04427  
; EARLIER FILING DATE: 1997-03-20  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 27  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; OTHER INFORMATION: Upstream primer for PCR amplification.  
US-08-940-544-1

Query Match 100.0%; Score 27; DB 8; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.011;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATCCT 27

Db 1 GCGGCGCAATTGAAGTTATGTATCCT 27

```
RESULT 2
US-10-448-256-6
; Sequence 6, Application US/10448256
; Publication No. US20040043401A1
; GENERAL INFORMATION:
; APPLICANT: Sadelain, Michel
; APPLICANT: Brentjens, Renier
; APPLICANT: Maher, John
; TITLE OF INVENTION: Chimeric T Cell Receptors
; FILE REFERENCE: MSK-P-058
; CURRENT APPLICATION NUMBER: US/10/448,256
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/383,872
; PRIOR FILING DATE: 2002-05-28
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 328
; TYPE: DNA
; ORGANISM: human
US-10-448-256-6

Query Match      70.4%; Score 19; DB 16; Length 328;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
   |||||
Db 4 AATTGAAGTTATGATCCT 22

RESULT 3
US-09-997-722-258
; Sequence 258, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 258
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-997-722-258

Query Match      70.4%; Score 19; DB 11; Length 663;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
   |||||
Db 339 AATTGAAGTTATGATCCT 357

RESULT 4
US-09-738-546-1
; Sequence 1, Application US/09738546
; Patent No. US20020006403A1
; GENERAL INFORMATION:
; APPLICANT: YU, XUE-ZHONG
; APPLICANT: ANASETTI, CLAUDIO
; TITLE OF INVENTION: CD28-SPECIFIC ANTIBODY COMPOSITIONS FOR USE IN METHODS
; OF IMMUNOSUPPRESSION
; FILE REFERENCE: FHCC:00705
```

```
; CURRENT APPLICATION NUMBER: US/09/738,546
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/170,857.
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-546-1

Query Match      70.4%; Score 19; DB 9; Length 1514;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
   |||||
Db 438 AATTGAAGTTATGATCCT 456

RESULT 5
US-09-836-544-7
; Sequence 7, Application US/09836544
; Publication No. US20040072283A1
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: Rapid Immunoselection Cloning Method
; FILE REFERENCE: 11-88L
; CURRENT APPLICATION NUMBER: US/09/836,544
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 07/983,647
; PRIOR FILING DATE: 1992-12-01
; PRIOR APPLICATION NUMBER: US 07/553,759
; PRIOR FILING DATE: 1990-07-13
; PRIOR APPLICATION NUMBER: US 07/498,809
; PRIOR FILING DATE: 1990-03-23
; PRIOR APPLICATION NUMBER: US 07/379,076
; PRIOR FILING DATE: 1989-07-13
; PRIOR APPLICATION NUMBER: US 07/160,416
; PRIOR FILING DATE: 1988-02-25
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (100)..(759)
US-09-836-544-7

Query Match      70.4%; Score 19; DB 11; Length 1514;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
   |||||
Db 438 AATTGAAGTTATGATCCT 456

RESULT 6
US-10-207-655-98
; Sequence 98, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden-Ledbetter, Martha S.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207,655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
```

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 98
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-207-655-98

Query Match          70.4%; Score 19; DB 15; Length 1514;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTCCT 27
   |||||
Db 438 AATTGAAGTTATGTCCT 456

RESULT 7
US-10-641-643-1316
; Sequence 1316, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; SUGAN G. STUART
; JEFFREY J. SEILHAMER
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/641,643
; FILING DATE: 14-Aug-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1316:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 9338444
; SEQUENCE DESCRIPTION: SEQ ID NO: 1316 :
US-10-641-643-1316

Query Match          70.4%; Score 19; DB 16; Length 1514;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTCCT 27
   |||||
Db 438 AATTGAAGTTATGTCCT 456
```

```
RESULT 8
US-10-076-934-1
; Sequence 1, Application US/10076934
; Publication No. US20030170232A1
; GENERAL INFORMATION:
; APPLICANT: O'Hara Jr., Richard
; APPLICANT: Nagelin, Ann Marie
; TITLE OF INVENTION: AGENTS THAT SPECIFICALLY BLOCK
; CD28-MEDIATED SIGNALING AND USES THEREFOR
; FILE REFERENCE: GNN-028
; CURRENT APPLICATION NUMBER: US/10/076,934
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 60/269,756
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3803
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-076-934-1

Query Match          70.4%; Score 19; DB 15; Length 3803;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTCCT 27
   |||||
Db 560 AATTGAAGTTATGTCCT 578

RESULT 9
US-09-997-722-257
; Sequence 257, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 257
; LENGTH: 3804
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-997-722-257

Query Match          70.4%; Score 19; DB 11; Length 3804;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGTCCT 27
   |||||
Db 561 AATTGAAGTTATGTCCT 579

RESULT 10
US-10-143-238-1
; Sequence 1, Application US/10143238
; Publication No. US20030166502A1
; GENERAL INFORMATION:
; APPLICANT: Green, Jonathan M.
; APPLICANT: Shaw, Andrew S.
; TITLE OF INVENTION: DIFFERENTIAL REGULATION OF T CELL SURVIVAL AND
; PROLIFERATION
```

Query Match	70.4%;	Score 19;	DB 15;	Length 3804;
Best Local Similarity	100.0%;	Pred. No. 1.4e+02;	Indels 0;	Gaps 0;
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	9	AATTGAAGTTATGTATCCT 27		
Db	561	AATTGAAGTTATGTATCCT 579		
RESULT 11				
US-09-835-297-3				
Sequence 3,	Application US/09835297			
Publication No.	US20030086932A1			
GENERAL INFORMATION:				
APPLICANT:	Bluestone, Jeffrey			
APPLICANT:	Collins, Mary			
APPLICANT:	Griffin, Mathew			
APPLICANT:	Kranz, David			
TITLE OF INVENTION:	SURFACE-BOUND ANTIGEN BINDING PORTIONS OF ANTIBODIES			
TITLE OF INVENTION:	THAT BIND TO CTLA4 AND USES THEREFOR			
FILE REFERENCE:	GNN-014CP			
CURRENT APPLICATION NUMBER:	US/09/835,297			
CURRENT FILING DATE:	2001-04-12			
PRIOR APPLICATION NUMBER:	60/196,851			
PRIOR FILING DATE:	2000-04-12			
NUMBER OF SEQ ID NOS:	4			
SOFTWARE:	Patentin Ver. 2.0			
SEQ ID NO 3				
LENGTH:	3806			
TYPE:	DNA			
ORGANISM:	Homo sapiens			
US-09-835-297-3				
Query Match	70.4%;	Score 19;	DB 10;	Length 3806;
Best Local Similarity	100.0%;	Pred. No. 1.4e+02;	Indels 0;	Gaps 0;
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	9	AATTGAAGTTATGTATCCT 27		
Db	561	AATTGAAGTTATGTATCCT 579		
RESULT 12				
US-09-997-722-256				
Sequence 256,	Application US/09997722			
Publication No.	US20040072154A1			
GENERAL INFORMATION:				
APPLICANT:	Morris, David			
APPLICANT:	Engelhardt, Eric			
TITLE OF INVENTION:	NOVEL COMPOSITIONS AND METHODS FOR CANCER			
FILE REFERENCE:	A-71171/RMS/DCF			
CURRENT APPLICATION NUMBER:	US/09/997,722			
CURRENT FILING DATE:	2001-11-30			
PRIOR APPLICATION NUMBER:	US 09/747,377			
PRIOR FILING DATE:	2000-12-22			
PRIOR APPLICATION NUMBER:	US 09/798,586			
PRIOR FILING DATE:	2001-03-02			

```
; ORGANISM: Streptococcus pneumoniae
US-10-097-111-10

Query Match      67.4%; Score 18.2; DB 15; Length 56506;
Best Local Similarity 87.0%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GCGGCAATTGAAGTTATGTATCC 26
    |||||
Db 38279 GCGGCAACTGAAGTCATGTATCC 38301

RESULT 15
US-10-437-963-96729
; Sequence 96729, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 96729
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1112)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_94799C.1
US-10-437-963-96729

Query Match      66.7%; Score 18; DB 17; Length 1112;
Best Local Similarity 80.8%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATCC 26
    |||||
Db 752 GCAGCGCATTCGAAGTGATATCC 777

RESULT 16
US-10-260-238-2623
; Sequence 2623, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
```

```
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 2623
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (1674)..(1674)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-2623

Query Match      66.7%; Score 18; DB 16; Length 2000;
Best Local Similarity 80.8%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTATCC 26
    |||||
Db 1484 GCAGCGCATTCGAAGTGATATCC 1509

RESULT 17
US-10-282-122A-21821
; Sequence 21821, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21821
; LENGTH: 2178
; TYPE: DNA
; ORGANISM: Enterococcus faecium
```

US-10-282-122A-21821

Query Match 65.9%; Score 17.8; DB 16; Length 2178;  
Best Local Similarity 90.5%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 GCAATTGAAGTTATGTATCCT 27  
||||| ||||| ||||| ||||| |||||  
Db 1042 GCAATCGAAGTTATTTATCCT 1062

RESULT 18

US-10-697-828-23/c  
; Sequence 23, Application US/10697828  
; Publication No. US20040185546A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen, Steven  
; APPLICANT: Lee, Jin Kyu  
; APPLICANT: Hemmerich, Stefan  
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6  
; FILE REFERENCE: UCAL-138DIV  
; CURRENT APPLICATION NUMBER: US/10/697,828  
; CURRENT FILING DATE: 2003-10-29  
; PRIOR APPLICATION NUMBER: 09/593,828  
; PRIOR FILING DATE: 2000-06-13  
; PRIOR APPLICATION NUMBER: 60/144,694  
; PRIOR FILING DATE: 1999-07-20  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 3624.  
; TYPE: DNA  
; ORGANISM: mouse  
US-10-697-828-23

Query Match 65.9%; Score 17.8; DB 17; Length 3624;  
Best Local Similarity 90.5%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 GCAATTGAAGTTATGTATCCT 27  
||||| ||||| ||||| ||||| |||||  
Db 2788 GCAAGTGAAGTTTGTATCCT 2768

RESULT 19

US-10-697-828-19/c  
; Sequence 19, Application US/10697828  
; Publication No. US20040185546A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen, Steven  
; APPLICANT: Lee, Jin Kyu  
; APPLICANT: Hemmerich, Stefan  
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6  
; FILE REFERENCE: UCAL-138DIV  
; CURRENT APPLICATION NUMBER: US/10/697,828  
; CURRENT FILING DATE: 2003-10-29  
; PRIOR APPLICATION NUMBER: 09/593,828  
; PRIOR FILING DATE: 2000-06-13  
; PRIOR APPLICATION NUMBER: 60/144,694  
; PRIOR FILING DATE: 1999-07-20  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 4121  
; TYPE: DNA  
; ORGANISM: mouse  
US-10-697-828-19

Query Match 65.9%; Score 17.8; DB 17; Length 4121;  
Best Local Similarity 90.5%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 GCAATTGAAGTTATGTATCCT 27

Db 2788 GCAAGTGAAGTTTGTATCCT 2768  
||||| ||||| ||||| ||||| |||||

RESULT 20

US-10-697-828-16/c  
; Sequence 16, Application US/10697828  
; Publication No. US20040185546A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen, Steven  
; APPLICANT: Lee, Jin Kyu  
; APPLICANT: Hemmerich, Stefan  
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6  
; FILE REFERENCE: UCAL-138DIV  
; CURRENT APPLICATION NUMBER: US/10/697,828  
; CURRENT FILING DATE: 2003-10-29  
; PRIOR APPLICATION NUMBER: 09/593,828  
; PRIOR FILING DATE: 2000-06-13  
; PRIOR APPLICATION NUMBER: 60/144,694  
; PRIOR FILING DATE: 1999-07-20  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 5084  
; TYPE: DNA  
; ORGANISM: mouse  
US-10-697-828-16

Query Match 65.9%; Score 17.8; DB 17; Length 5084;  
Best Local Similarity 90.5%; Pred. No. 5.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 GCAATTGAAGTTATGTATCCT 27  
||||| ||||| ||||| ||||| |||||  
Db 3288 GCAAGTGAAGTTTGTATCCT 3268

RESULT 21

US-10-087-192-25  
; Sequence 25, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 48445  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(48445)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-25

Query Match 65.9%; Score 17.8; DB 13; Length 48445;  
Best Local Similarity 90.5%; Pred. No. 7.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GCCGCAATTGAAGTTATGTAT 24  
||||| ||||| ||||| ||||| |||||  
Db 47072 GCCTAAATTGAAGTTATGTAT 47092

```
RESULT 22
US-09-966-277-21
; Sequence 21, Application US/09966277
; Patent No. US20020169306A1
; GENERAL INFORMATION:
; APPLICANT: KITAZATO, Kaio
; APPLICANT: SHU, Tsugumine
; APPLICANT: KUMA, Hidekazu
; APPLICANT: UEDA, Yasuji
; APPLICANT: ASAKAWA, Makoto
; APPLICANT: HASEGAWA, Mamoru
; APPLICANT: IIDA, Akihiro
; APPLICANT: TOKITOU, Fumino
; APPLICANT: HIRATA, Takahiro
; APPLICANT: TOKUSUMI, Tsuyoshi
; APPLICANT: INOUE, Makoto
; TITLE OF INVENTION: ENVELOPE GENE-DEFICIENT PARAMYXOVIRUS
; TITLE OF INVENTION: VECTOR
; FILE REFERENCE: 50026/028001
; CURRENT APPLICATION NUMBER: US/09/966,277
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/JP00/03195
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: JP 2001/283451
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP 11/200739
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially Synthesized Sequence
US-09-966-277-21

Query Match      65.2%; Score 17.6; DB 9; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GCGGCCGCAATTGAAGTTATGTAT 24
          |||||
Db      15 GCGGCCGCAATGCGAGATATCTAT 38

RESULT 23
US-09-966-277-22/c
; Sequence 22, Application US/09966277
; Patent No. US20020169306A1
; GENERAL INFORMATION:
; APPLICANT: KITAZATO, Kaio
; APPLICANT: SHU, Tsugumine
; APPLICANT: KUMA, Hidekazu
; APPLICANT: UEDA, Yasuji
; APPLICANT: ASAKAWA, Makoto
; APPLICANT: HASEGAWA, Mamoru
; APPLICANT: IIDA, Akihiro
; APPLICANT: TOKITOU, Fumino
; APPLICANT: HIRATA, Takahiro
; APPLICANT: TOKUSUMI, Tsuyoshi
; APPLICANT: INOUE, Makoto
; TITLE OF INVENTION: ENVELOPE GENE-DEFICIENT PARAMYXOVIRUS
; TITLE OF INVENTION: VECTOR
; FILE REFERENCE: 50026/028001
; CURRENT APPLICATION NUMBER: US/09/966,277
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/JP00/03195
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: JP 2001/283451
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP 11/200739
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially Synthesized Sequence
US-09-966-277-21

Query Match      65.2%; Score 17.6; DB 9; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GCGGCCGCAATTGAAGTTATGTAT 24
          |||||
Db      15 GCGGCCGCAATGCGAGATATCTAT 38
```

```
RESULT 24
US-09-966-930-21
; Sequence 21, Application US/09966930
; Publication No. US20030022376A1
; GENERAL INFORMATION:
; APPLICANT: KITAZATO, Kaio
; APPLICANT: SHU, Tsugumine
; APPLICANT: KUMA, Hidekazu
; APPLICANT: UEDA, Yasuji
; APPLICANT: ASAKAWA, Makoto
; APPLICANT: HASEGAWA, Mamoru
; APPLICANT: IIDA, Akihiro
; APPLICANT: HIRATA, Takahiro
; APPLICANT: INOUE, Makoto
; TITLE OF INVENTION: PARAMYXOVIRUS-DERIVED RNP
; FILE REFERENCE: 50026/029001
; CURRENT APPLICATION NUMBER: US/09/966,930
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/JP00/03194
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: JP 2001/283451
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP 11/200740
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially Synthesized Sequence
US-09-966-930-21

Query Match      65.2%; Score 17.6; DB 10; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GCGGCCGCAATTGAAGTTATGTAT 24
          |||||
Db      15 GCGGCCGCAATGCGAGATATCTAT 38

RESULT 25
US-09-966-930-22/c
; Sequence 22, Application US/09966930
; Publication No. US20030022376A1
; GENERAL INFORMATION:
; APPLICANT: KITAZATO, Kaio
; APPLICANT: SHU, Tsugumine
; APPLICANT: KUMA, Hidekazu
; APPLICANT: UEDA, Yasuji
; APPLICANT: ASAKAWA, Makoto
; APPLICANT: ASAKAWA, Makoto
```

```
; APPLICANT: HASEGAWA, Mamoru
; APPLICANT: IIDA, Akihiro
; APPLICANT: HIRATA, Takahiro
; APPLICANT: INOUE, Makoto
; TITLE OF INVENTION: PARAMYXOVIRUS-DERIVED RNP
; FILE REFERENCE: 50026/029001
; CURRENT APPLICATION NUMBER: US/09/966,930
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/JP00/03194
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: JP 2001/283451
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP 11/200740
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially Synthesized Sequence
US-09-966-930-22

Query Match      65.2%; Score 17.6; DB 10; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGCAAGTTATGTAT 24
Db 26 GCGGCGCAATGGCAGATATCTAT 3

RESULT 26
US-10-316-538-21
; Sequence 21, Application US/10316538
; Publication No. US20030166252A1
; GENERAL INFORMATION:
; APPLICANT: KITAZATO, Kaio
; APPLICANT: SHU, Tsugumine
; APPLICANT: KUMA, Hidekazu
; APPLICANT: UEDA, Yasuji
; APPLICANT: ASAKAWA, Makoto
; APPLICANT: HASEGAWA, Mamoru
; APPLICANT: IIDA, Akihiro
; APPLICANT: HIRATA, Takahiro
; APPLICANT: INOUE, Makoto
; APPLICANT: TOKUSUMI, Yumiko
; TITLE OF INVENTION: PARAMYXOVIRUS-DERIVED RNP
; FILE REFERENCE: D3-102PCT-USC1
; CURRENT APPLICATION NUMBER: US/10/316,538
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/JP02/09558
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 09/966,930
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: JP 2001-283451
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: PCT/JP00/03194
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: JP 1999-200740
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: artificially
; OTHER INFORMATION: synthesized sequence
US-10-316-538-22

Query Match      65.2%; Score 17.6; DB 15; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGCAAGTTATGTAT 24
Db 26 GCGGCGCAATGGCAGATATCTAT 3

RESULT 28
US-10-316-535-21
; Sequence 21, Application US/10316535
; Publication No. US20030170266A1
; GENERAL INFORMATION:
; APPLICANT: KITAZATO, Kaio
; APPLICANT: SHU, Tsugumine
; APPLICANT: KUMA, Hidekazu
; APPLICANT: UEDA, Yasuji
; APPLICANT: ASAKAWA, Makoto
; APPLICANT: HASEGAWA, Mamoru
; APPLICANT: IIDA, Akihiro
; APPLICANT: TOKITOU, Fumino
; APPLICANT: HIRATA, Takahiro
; APPLICANT: TOKUSUMI, Tsuyoshi
```



```
/ APPLICANT: INOUE, Makoto
/ APPLICANT: TOKUSUMI, Yumiko
/ TITLE OF INVENTION: ENVELOPE GENE-DEFICIENT PARAMYXOVIRUS VECTOR
/ FILE REFERENCE: D3-103PCT-USC1
/ CURRENT APPLICATION NUMBER: US/10/316,535
/ PRIOR FILING DATE: 2002-12-10
/ PRIOR APPLICATION NUMBER: PCT/JP02/09558
/ PRIOR FILING DATE: 2002-09-18
/ PRIOR FILING DATE: 2002-09-18
/ PRIOR APPLICATION NUMBER: US 09/966,277
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: JP 2001-283451
/ PRIOR FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: PCT/JP00/03195
/ PRIOR FILING DATE: 2000-05-18
/ PRIOR APPLICATION NUMBER: JP 1999-200739
/ PRIOR FILING DATE: 1999-05-18
/ NUMBER OF SEQ ID NOS: 68
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 21
/ LENGTH: 40
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: artificially
/ OTHER INFORMATION: synthesized sequence
US-10-316-535-21

Query Match      65.2%; Score 17.6; DB 15; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTAT 24
    |||||
Db 15 GCGGCGCAATGCGAGATATCTAT 38

RESULT 29
US-10-316-535-22/c
/ Sequence 22, Application US/10316535
/ Publication No. US20030170266A1
/ GENERAL INFORMATION:
/ APPLICANT: KITAZATO, Kaio
/ APPLICANT: SHU, Tsugumine
/ APPLICANT: KUMA, Hidekazu
/ APPLICANT: UEDA, Yasuji
/ APPLICANT: ASAKAWA, Makoto
/ APPLICANT: HASEGAWA, Mamoru
/ APPLICANT: IIDA, Akihiro
/ APPLICANT: TOKITOU, Fumino
/ APPLICANT: HIRATA, Takahiro
/ APPLICANT: TOKUSUMI, Tsuyoshi
/ APPLICANT: INOUE, Makoto
/ APPLICANT: TOKUSUMI, Yumiko
/ TITLE OF INVENTION: ENVELOPE GENE-DEFICIENT PARAMYXOVIRUS VECTOR
/ FILE REFERENCE: D3-103PCT-USC1
/ CURRENT APPLICATION NUMBER: US/10/316,535
/ PRIOR FILING DATE: 2002-12-10
/ PRIOR APPLICATION NUMBER: PCT/JP02/09558
/ PRIOR FILING DATE: 2002-09-18
/ PRIOR APPLICATION NUMBER: US 09/966,277
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: JP 2001-283451
/ PRIOR FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: PCT/JP00/03195
/ PRIOR FILING DATE: 2000-05-18
/ PRIOR APPLICATION NUMBER: JP 1999-200739
/ PRIOR FILING DATE: 1999-05-18
/ NUMBER OF SEQ ID NOS: 68
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 22
/ LENGTH: 40
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
```

```
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: artificially
/ OTHER INFORMATION: synthesized sequence
US-10-316-535-22

Query Match      65.2%; Score 17.6; DB 15; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTAT 24
    |||||
Db 26 GCGGCGCAATGCGAGATATCTAT 3

RESULT 30
US-10-784-710-3
/ Sequence 3, Application US/10784710
/ Publication No. US20040137627A1
/ GENERAL INFORMATION:
/ APPLICANT: Tokusumi, Tsuyoshi
/ APPLICANT: Iida, Akihiro
/ APPLICANT: Hasegawa, Mamoru
/ APPLICANT: Nagai, Yoshiyuki
/ TITLE OF INVENTION: Paramyxovirus vectors used for transfer
/ TITLE OF INVENTION: of foreign genes
/ FILE REFERENCE: 50026/025001
/ CURRENT APPLICATION NUMBER: US/10/784,710
/ CURRENT FILING DATE: 2004-02-23
/ PRIOR APPLICATION NUMBER: US/09/702,498A
/ PRIOR FILING DATE: 2000-10-31
/ PRIOR APPLICATION NUMBER: JP 2000-152726
/ PRIOR FILING DATE: 2000-05-18
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 40
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Artificially synthesized sequence derived from the
/ OTHER INFORMATION: Sendai virus
US-10-784-710-3

Query Match      65.2%; Score 17.6; DB 17; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGGCGCAATTGAAGTTATGTAT 24
    |||||
Db 15 GCGGCGCAATGCGAGATATCTAT 38

RESULT 31
US-10-784-710-4/c
/ Sequence 4, Application US/10784710
/ Publication No. US20040137627A1
/ GENERAL INFORMATION:
/ APPLICANT: Tokusumi, Tsuyoshi
/ APPLICANT: Iida, Akihiro
/ APPLICANT: Hasegawa, Mamoru
/ APPLICANT: Nagai, Yoshiyuki
/ TITLE OF INVENTION: Paramyxovirus vectors used for transfer
/ TITLE OF INVENTION: of foreign genes
/ FILE REFERENCE: 50026/025001
/ CURRENT APPLICATION NUMBER: US/10/784,710
/ CURRENT FILING DATE: 2004-02-23
/ PRIOR APPLICATION NUMBER: US/09/702,498A
/ PRIOR FILING DATE: 2000-10-31
/ PRIOR APPLICATION NUMBER: JP 2000-152726
/ PRIOR FILING DATE: 2000-05-18
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 4
```

```
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially synthesized sequence derived from the
US-10-784-710-4

Query Match      65.2%; Score 17.6; DB 17; Length 40;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGCGCCCAATTGAAGTTATGTAT 24
    |||||
Db 26 GCGCGCCCAATGGCAGATATCTAT 3

RESULT 32
US-10-437-963-33246/c
; Sequence 33246, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 33246
; LENGTH: 2006
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2006)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37377C.1
US-10-437-963-33246

Query Match      65.2%; Score 17.6; DB 17; Length 2006;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GCGCGCAATTGAAGTTATGTATCCT 27
    |||||
Db 1871 GCGCGCAATTGAAGTTCTGAAACCT 1848

RESULT 33
US-10-264-049-404
; Sequence 404, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133p1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
```

```
; SEQ ID NO 404
; LENGTH: 2541
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-049-404

Query Match      65.2%; Score 17.6; DB 16; Length 2541;
Best Local Similarity 83.3%; Pred. No. 5.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GCGCGCAATTGAAGTTATGTATCCT 27
    |||||
Db 2430 GCGCGCAATTGAATTTATGTATCCT 2453

RESULT 34
US-10-057-726-16/c
; Sequence 16, Application US/10057726
; Publication No. US20030017549A1
; GENERAL INFORMATION:
; APPLICANT: Owens, Gary K.
; APPLICANT: Manabe, Ichiro
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR EXPRESSING POLYNUCLEOTIDES
; FILE REFERENCE: 021258-000200US
; CURRENT APPLICATION NUMBER: US/10/057,726
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: US 60/263,811
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 09/600,319
; PRIOR FILING DATE: 2000-07-13
; PRIOR APPLICATION NUMBER: WO PCT/US99/01038
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: US 60/071,300
; PRIOR FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 16011
; TYPE: DNA
; ORGANISM: Rattus sp.
US-10-057-726-16

Query Match      65.2%; Score 17.6; DB 14; Length 16011;
Best Local Similarity 83.3%; Pred. No. 7.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GCGCGCAATTGAAGTTATGTATCCT 27
    |||||
Db 6333 GCGCGCAATTGAAGTTATGTATCCT 6310

RESULT 35
US-09-769-736-186
; Sequence 186, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 186
; LENGTH: 41
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer
US-09-769-736-186

Query Match      64.4%; Score 17.4; DB 10; Length 41;
Best Local Similarity 77.8%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATCCT 27
    |||||
Db 3 GCGCGCGCAATTAGTGTATTCTCCT 29

RESULT 36
US-09-987-899-5599
; Sequence 5599, Application US/09987899
; Publication No. US20040116682A1
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; APPLICANT: Miller, Phillip W.
; APPLICANT: O Connell, Keith M.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; TITLE OF INVENTION: With the Carbon Assimilation Pathway
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US/09/987,899
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5599
; LENGTH: 154
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701169557H1
US-09-987-899-5599

Query Match      64.4%; Score 17.4; DB 11; Length 154;
Best Local Similarity 77.8%; Pred. No. 4.5e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATCCT 27
    |||||
Db 87 GAGCGCGCAATTCAGCTATCAATGCT 113

RESULT 37
US-09-987-899-5598
; Sequence 5598, Application US/09987899
; Publication No. US20040116682A1
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; APPLICANT: Miller, Phillip W.
; APPLICANT: O Connell, Keith M.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; TITLE OF INVENTION: With the Carbon Assimilation Pathway
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US/09/987,899
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5598
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Human

; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700243862H1
US-09-987-899-5598

Query Match      64.4%; Score 17.4; DB 11; Length 285;
Best Local Similarity 77.8%; Pred. No. 5e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATCCT 27
    |||||
Db 252 GAGCGCGCAATTCAGCTATCAATGCT 278

RESULT 38
US-10-242-535A-1095/c
; Sequence 1095, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1095
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-1095

Query Match      64.4%; Score 17.4; DB 16; Length 292;
Best Local Similarity 77.8%; Pred. No. 5e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATCCT 27
    |||||
Db 285 GCTGGTGCACTTAAAGTTATTTATCCT 259

RESULT 39
US-10-085-783A-1095/c
; Sequence 1095, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1095
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Human
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 1611.95 Seconds  
(without alignments)  
610.363 Million cell updates/sec

Title: US-09-786-502A-7  
Perfect score: 27  
Sequence: 1 gggcgcaattgaagtattatct 27

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.6	72.6	866	2	BF213303 601845014
C 2	19.2	71.1	852	5	BQ734454 AGENCOURT
C 3	19	70.4	304	5	BQ319411 MRO-CT045
C 4	19	70.4	446	4	BQ31520 BJ031520
C 5	19	70.4	450	5	BX101593 BX101593
C 6	19	70.4	510	2	AW949936 EST362006
C 7	19	70.4	596	6	CD691494 EST8017 h
C 8	19	70.4	611	9	AY402331 Pan trogl
C 9	19	70.4	654	5	BX955995 DKFZp781F
C 10	19	70.4	663	9	AY402330 Homo sapi
C 11	19	70.4	710	5	BX854672 BX854672
C 12	19	70.4	723	9	AG281788 Mus muscu
C 13	19	70.4	730	2	AW954205 EST366275
C 14	19	70.4	886	7	CK800588 AGENCOURT
C 15	19	70.4	935	4	BI950426 BVSME1002
C 16	19	70.4	1000	5	BX403908 BX403908
C 17	18.6	68.9	521	6	CB828552 LjNEST87h
C 18	18.6	68.9	687	8	BZ320247 hz06d07.9
C 19	18.6	68.9	815	6	CB901075 trico25xf
C 20	18.6	68.9	815	7	BF870869 trico25xf
C 21	18.6	68.9	818	8	BH463620 BOHQC78TF
C 22	18.6	68.9	831	5	BP512240 BP512240
C 23	18.6	68.9	919	9	CG303571 OGLA086TV
C 24	18.4	68.1	486	9	AL609371 Anopheles

25	18.4	68.1	628	8	AZ338890
C 26	18.4	68.1	720	7	N26591
C 27	18.4	68.1	999	9	CR143861
C 28	18.2	67.4	768	5	BU112231
C 29	18.2	67.4	788	5	BU335623
C 30	18.2	67.4	825	5	BU381772
C 31	18.2	67.4	838	9	CNS02N5T
C 32	18.2	67.4	1329	9	CL023772
C 33	18	66.7	205	5	BY592030
C 34	18	66.7	508	6	CB676984
C 35	18	66.7	561	8	AQ654959
C 36	18	66.7	562	5	BQ135922
C 37	18	66.7	599	8	AQ644499
C 38	18	66.7	621	8	AQ787083
C 39	18	66.7	625	2	BF649189
C 40	18	66.7	719	6	CB651428
C 41	18	66.7	724	8	CC313680
C 42	18	66.7	777	2	BE572142
C 43	18	66.7	783	7	CF924217
C 44	18	66.7	787	8	BH397370
C 45	18	66.7	798	6	CD781415

## ALIGNMENTS

RESULT 1  
BF213303/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF213303 866 bp mRNA linear EST 06-NOV-2000  
601845014F1 NIH\_MGC\_55 Homo sapiens cDNA clone IMAGE:4070522 5',  
mRNA sequence.  
BF213303  
BF213303.1 GI:11106889  
EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC Institute of Health, Mammalian Gene Collection (MGC)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM914 row: k column: 03  
High quality sequence stop: 582.

FEATURES  
source

1. .866  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4070522"  
/tissue\_type="from acute myelogenous leukemia"  
/lab\_host="DH10B (Tl phage-resistant)"  
/clone\_lib="NIH\_MGC\_55"  
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);  
Site 1: SfII (ggcgccctcgcc); Site 2: SfII  
(ggcattatggcc); Double-stranded cDNA was prepared from  
cell line RNA. 5' and 3' adaptors were used in cloning as  
follows: 5' adaptor sequence: 5'-CAGGCCATTATGCC-3' and  
3' adaptor sequence:  
5'-ATTCTAGCCAGCGGCCGCGACATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size  
1.65 kb (range 0.9-4.0 kb). 14/15 colonies contained  
inserts by PCR. This library was enriched for full-length  
clones and was constructed by Clontech Laboratories (Palo



COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshinigenes.nig.ac.jp  
The information of this clone is available through the following  
URL.  
http://xenopus.nibb.ac.jp.

FEATURES source  
Location/Qualifiers  
1. 446  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL007111"  
/tissue type="whole embryo"  
/dev\_stage="stage 15"  
/clone\_lib="NIBB Mochii normalized Xenopus neurula  
library"

ORIGIN  
Query Match 70.4%; Score 19; DB 4; Length 446;  
Best Local Similarity 81.5%; Pred. No. 2.8e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 GCGCGCCGAATGAAGTTATGTATCCT 27  
|||||  
Db 323 GCTGCAGAAATGAAGTTATGTATCCT 297

RESULT 5  
BX101593  
LOCUS  
DEFINITION BX101593 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone  
IMAGE:2048836; IMAGE:2048836, mRNA sequence.  
ACCESSION BX101593  
VERSION BX101593.1 GI:27844973  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 450)  
AUTHORS Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,  
Radelof, U., Schneider, D. and Korn, B.  
TITLE Human Unigeneset - RZPD3  
JOURNAL Unpublished (2003)  
COMMENT Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD; IMAGp998N055044.  
RZPDLIB: I.M.A.G.E. cDNA Clone Collection;  
http://www.rzpd.de/CloneCards/cgi-  
bin/showLib.pl.cgi?responderLibNo=972 Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available royalty-free from RZPD;  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.

FEATURES source  
Location/Qualifiers  
1. 450  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGp998N055044 ; IMAGE:2048836"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/clone\_lib="Soares\_total\_fetus\_Nb2HF8\_9w"

COMMENT /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from pooled 8-9 week  
(total) fetus material with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTITTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN  
Query Match 70.4%; Score 19; DB 5; Length 450;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 AATTGAAGTTATGTATCCT 27  
|||||  
Db 418 AATTGAAGTTATGTATCCT 436

RESULT 6  
AW949936  
LOCUS  
DEFINITION EST162006 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW949936  
VERSION AW949936.1 GI:8139576  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 510)  
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C.,  
Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and  
Quackenbush, J.  
TITLE Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
JOURNAL Unpublished (2000)  
COMMENT Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 8  
Seq primer: Reverse.  
Location/Qualifiers  
1. 510  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGE"  
/note="Vector: pBluescriptSkm"

ORIGIN  
Query Match 70.4%; Score 19; DB 2; Length 510;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 AATTGAAGTTATGTATCCT 27  
|||||  
Db 337 AATTGAAGTTATGTATCCT 355

RESULT 7  
CD691494  
LOCUS  
DEFINITION EST8017 human nasopharynx Homo sapiens cDNA, mRNA sequence.  
ACCESSION CD691494  
VERSION CD691494.1 GI:32213267  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Liu,X.-O., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and
Zeng,Y.-X.
TITLE Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL Unpublished (2003)
COMMENT Contact: Yixin Zeng
Cancer Center
Sun Yat-sen University
651 Dongfeng Road East, Guangzhou 510060, China
Tel: 86-1380-9770-743
Fax: 86-20-8775-4506
Email: yxzeng@gzsums.edu.cn.
FEATURES
source
1..596
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/notes="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"
ORIGIN
Query Match 70.4%; Score 19; DB 6; Length 596;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
|||||
Db 465 AATTGAAGTTATGATCCT 483

RESULT 8
AY402331
LOCUS AY402331 611 bp DNA linear GSS 15-DEC-2003
DEFINITION Pan troglodytes CD28 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY402331
VERSION AY402331.1 GI:39758317
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 611)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 611)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..611
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>611
/gene="CD28"

QY 9 AATTGAAGTTATGATCCT 27
|||||
Db 400 AATTGAAGTTATGATCCT 418

RESULT 10
AY402330
LOCUS AY402330 663 bp DNA linear GSS 15-DEC-2003
DEFINITION Homo sapiens CD28 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY402330
VERSION AY402330.1 GI:39758316

```

```

/locus_tag="HCM1188"

Query Match 70.4%; Score 19; DB 9; Length 611;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AATTGAAGTTATGATCCT 27
|||||
Db 287 AATTGAAGTTATGATCCT 305

RESULT 9
BX955995
LOCUS BX955995 654 bp mRNA linear EST 01-MAR-2004
DEFINITION DKFZp781F1075 r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
DKFZp781F1075 5', mRNA sequence.
ACCESSION BX955995
VERSION BX955995.1 GI:43439640
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 654)
AUTHORS Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
Newes,H.W., Well,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
TITLE EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
Wellenreuther,R., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp781F1075) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1..654
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp781F1075"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="781 (synonym: hlcc4)"
/notes="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

QY 9 AATTGAAGTTATGATCCT 27
|||||
Db 400 AATTGAAGTTATGATCCT 418

RESULT 10
AY402330
LOCUS AY402330 663 bp DNA linear GSS 15-DEC-2003
DEFINITION Homo sapiens CD28 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY402330
VERSION AY402330.1 GI:39758316

```



KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 663)  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 663)  
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES Location/Qualifiers  
source  
1..663  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
gene  
1..663  
/gene="CD28"  
/locus\_tag="HCM1188"  
ORIGIN  
Query Match 70.4%; Score 19; DB 9; Length 663;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 AATTGAAGTTATGATCCT 27  
1 |||||  
Db 339 AATTGAAGTTATGATCCT 357  
|||||  
BX854672 710 bp mRNA linear EST 11-DEC-2003  
BX854672 Kirschner embryo St10.14 Xenopus laevis cDNA clone  
IMAGp998F228575; IMAGE:3516309 5', mRNA sequence.  
ACCESSION BX854672 GI:39745925  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Xenopus laevis (African clawed frog)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Xenopus.  
1 (bases 1 to 710)  
Heil, O., Neubert, P., Peters, M., Radelof, U., Schneider, D.,  
Schroth, A., Korn, B. and Landgrebe, J.  
Xenopus laevis UniGene Set 1 (RZPDLIB No.988)  
Xenopus laevis UniGene Set 1 (RZPDLIB No.988)  
TITLE  
JOURNAL  
COMMENT  
Contact: Ina Rolf  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD: IMAGp998F228575.  
RZPDLIB: I.M.A.G.E. cDNA Clone Collection (amp- resistant) (RZPDLIB  
No.998) http://www.rzpd.de/cgi-  
bin/products/showlib.pl.cgi/response?libNo=998 RZPDLIB: Xenopus  
laevis UniGene Set 1 (RZPDLIB No.988) http://www.rzpd.de/cgi-  
bin/products/showlib.pl.cgi/response?libNo=988 Contact: Ina Rolf  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101

Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available royalty-free from RZPD;  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
SP6, 5' ATTAGTGACACTAG 3'.  
FEATURES Location/Qualifiers  
source  
1..710  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="IMAGp998F228575; IMAGE:3516309"  
/tissue type="pooled embryos (stage 10-14)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="Kirschner embryo St10.14"  
/note="Vector: pCS2+; Site1: NotI; Site2: SalI;  
Size-selected for average insert size 1.2 kb. Library was  
constructed and donated by M. Kirschner (Harvard Medical  
School)."  
ORIGIN  
Query Match 70.4%; Score 19; DB 5; Length 710;  
Best Local Similarity 81.5%; Pred. No. 2.9e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 GCGCGCAATTGAAGTTATGATCCT 27  
1 |||||  
Db 353 GCTGCAGAAATGAAGTTATGATCCT 327  
|||||  
RESULT 12  
AG281788 723 bp DNA linear GSS 02-JUN-2004  
LOCUS Mus musculus molossinus DNA, clone:MSMg01-053J22.TU, genomic survey  
DEFINITION  
ACCESSION AG281788  
VERSION AG281788.1 GI:47854665  
KEYWORDS GSS.  
SOURCE Mus musculus molossinus  
ORGANISM Mus musculus molossinus  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
BAC end Sequences of Library MSMg01  
Unpublished  
2 (bases 1 to 723)  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
Direct Submission  
TITLE  
JOURNAL  
COMMENT  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
Tsukuba Institute, Bio Resource Center, Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : TJ  
LIBRARY  
Vector : pBACe3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.  
FEATURES Location/Qualifiers  
source  
1..723  
/organism="Mus musculus molossinus"  
/mol\_type="Genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-053J22.TU"



Fax: 864 656 4293  
Email: [twing@clemons.edu](mailto:twing@clemons.edu)  
Total hg bases = 369  
Seq primer: AATTACCTCTACTAAGGG  
High quality sequence stop: 913.

## FEATURES

source

```
1. .935
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSME10020K12f"
/tissue_type="Spike"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare spike EST library HVCDNA0012
(Fusarium infected)"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI. Plants were grown at the University of Minnesota in
the GJ Muehlbauer lab; spikes were harvested and snap
frozen at 0, 1, 2, 3, 4, 5, 6, and 8 days after Fusarium
graminearum inoculation (Heinen). In the TJ Close lab at
the University of California, Riverside, total RNA was
prepared from each sample pool, equal quantities of all
eight RNA pools were combined, poly(A) RNA was purified
from the mixture, one primary unamplified cDNA library was
made, and 1 million pfu were in vivo excised to give
pBluescript SK(-) cDNA phagemids (Choi, Fenton,
Malatras). Phagemids were plated and picked at the
Clemson University Genomics Institute (CUGI) (Begum,
Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
Oates, Rambo, Main). The sequence has been trimmed to
remove vector sequence and contains a minimum of 100 bases
of phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders/Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
```

## ORIGIN

```
Query Match 70.4%; Score 19; DB 4; Length 935;
Best Local Similarity 81.5%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 GCGCGCGCAATTGAAGTTATGTATCCT 27
||||| | | | | | | | | | | | | | | | | | | | |
Db 78 GCGCGCAGAGAGTGAAGTTATGTATCCT 52
```

## RESULT 16

```
EX403908/c 1000 bp mRNA linear EST 03-MAY-2004
EX403908 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CL0BB005ZE06 5-PRIME, mRNA sequence.
```

```
EX403908
EX403908.2 GI:46957179
EST.
```

## ORGANISM

```
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
1. (bases 1 to 1000)
```

```
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
```

```
Full-length cDNA libraries and normalization
```

```
Unpublished (2001)
```

```
On May 15, 2003 this sequence version replaced gi:30768436.
```

```
Contact: Genoscope
```

```
Genoscope - Centre National de Sequencage
```

BP 191 91006 EVRY cedex - France  
Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with NotI and cloned  
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
This sequence belongs to sequence cluster 6542.f  
For more information about this cluster, see  
<http://www.genoscope.cns.fr/cdna?e=CL0BB005ZE06RPL&c=6542.f>.

## FEATURES

source

```
1. .1000
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL0BB005ZE06"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with NotI and cloned into
the NotI and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
```

## ORIGIN

```
Query Match 70.4%; Score 19; DB 5; Length 1000;
Best Local Similarity 74.1%; Pred. No. 3.1e+02;
Matches 20; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 GCGCGCGCAATTGAAGTTATGTATCCT 27
||||| | | | | | | | | | | | | | | | | | | | |
```

```
Db 987 GCKGCGGCMDDTCCAGTGATGTACCT 961
||||| | | | | | | | | | | | | | | | | | | | |
```

## RESULT 17

```
CB828552
LOCUS 521 bp mRNA linear EST 16-MAY-2003
LjNEST87hLr Lotus japonicus nodule library 5 and 7 week-old Lotus
corniculatus var. japonicus cDNA 5', mRNA sequence.
```

```
CB828552
CB828552.1 GI:29968590
```

## KEYWORDS

```
EST.
Lotus corniculatus var. japonicus (Lotus japonicus)
```

## ORGANISM

```
Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
Lotus.
```

```
1 (bases 1 to 521)
```

```
Colebatch, G., Freund, S., Trevasakis, B and Udvardi, M.
```

```
Lotus japonicus root nodule ESTs: tools for functional genomics
```

```
Unpublished (2000)
```

```
Contact: Udvardi MK
```

```
Molecular Plant Nutrition
```

```
Max Planck Institute of Molecular Plant Physiology
```

```
Am Muehlenberg 1, 14476 Golm, Germany
```

```
Fax: 49 331 567 8250
```

```
Email: udvardi@mpimp-golm.mpg.de
```

```
Seq primer: T7
```

```
High quality sequence stop: 521.
```

```
Location/Qualifiers
```

```
1. .521
```

```
/organism="Lotus corniculatus var. japonicus"
```

```
/mol_type="mRNA"
```

```
/cultivar="Gifu (B-129)"
```

```
/db_xref="taxon:34305"
```

```
/dev_stage="5 and 7 week-old plants"
```

```
/clone_lib="Lotus japonicus nodule library 5 and 7
```

```
week-old"
```

```
/note="Organ: Nodule; Vector: pSPORT1; Site 1: SalI;
```

```
Site 2: NotI; The library was prepared using mRNA
```

```
extracted from nodules of 5 and 7 week-old Lotus plants.
```

```
Nodules were induced by, and contained Mesorhizobium
```

```

ORIGIN
Query Match      68.9%; Score 18.6; DB 6; Length 521;
Best Local Similarity 84.0%; Pred. No. 4.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25
    ||||| ||||| ||||| ||||| |||||
Db 220 GCGGTCGCACTTGCGAGTTGTATC 244

RESULT 18
BZ320247
LOCUS
DEFINITION      BZ320247 687 bp DNA linear GSS 06-NOV-2002
                  hz06d07.g1 WGS-ZmaySF (JM107 adapted methyl filtered) Zea mays
                  genomic clone hz06d07 5', genomic survey sequence.
ACCESSION      BZ320247
VERSION        BZ320247
KEYWORDS       BZ320247.1 GI:24695092
SOURCE         GSS.
ORGANISM       Zea mays
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
                Clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 687)
AUTHORS        Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,
                Katzenburger, F., King, L., Miller, B., Muller, S., Martienssen, R.A.,
                Zutavern, T., McCombie, W.R. and Martienssen, L.A.
                Genomic shotgun sequences from Zea mays (methyl-filtered)
                Unpublished (2002)
                Contact: W. Richard McCombie
                Lita Annenberg Hazen Genome Sequencing Center
                Cold Spring Harbor Laboratory
                PO Box 100, Cold Spring Harbor, NY 11724, USA
                Tel.: 516 367 8884
                Fax: 516 367 8874
                Email: mcombie@cshl.org
                Plate: hz06 row: d column: 07
                Seq primer: -21M13UnivRev
                Class: shotgun
                High quality sequence stop: 687.
FEATURES       source
                Location/Qualifiers
                1..687
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /cultivar="B73"
                /db_xref="taxon:4577"
                /clone="hz06d07"
                /lab_host="JM107 or DH5a"
                /clone_lib="WGS-ZmaySF (JM107 adapted methyl filtered)"
                /notes="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
                The vector was digested with XbaI and one nucleotide was
                added by fill in in the recessive 3' end. The genomic DNA
                was nebulized, end repaired, adaptor ligated and size
                fractionated using sephadex. The resulting fragments were
                between 0.8 and 3 kb and were cloned into the vector
                (-x/y reads in M13mp19, .b/g reads in pUC19). The same
                ligation was transformed in either JM107 or DH5a."
ORIGIN
Query Match      68.9%; Score 18.6; DB 8; Length 687;
Best Local Similarity 84.0%; Pred. No. 4.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GCGCGCAATTGAAGTTATGTATCCT 27
    ||||| ||||| ||||| ||||| |||||
Db 55 GCGCACATATGAAGTTATGTATCCT 79

RESULT 19
CB901075/c
LOCUS
DEFINITION      trico25xf21 T.reesei mycelial culture, Version 3 april Hypocrea
                  jecorina cDNA clone trico25xf21, mRNA sequence.
ACCESSION      CB901075
VERSION        CB901075.1 GI:30115733
KEYWORDS       EST.
SOURCE         Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM       Hypocrea jecorina
                Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
                Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE      Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
                Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
                Kelley, A.S., Meerman, H.J., Mitchell, F., Mitchinson, C.,
                Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
                Transcriptional regulation of biomass-degrading enzymes in the
                filamentous fungus Trichoderma reesei
                J. Biol. Chem. 278 (34), 31988-31997 (2003)
                22803314
                12789920
                Contact: Pamela K. Foreman
                Genencor Intl.
                925 Page Mill Road, Palo Alto, CA 94304, USA
                Tel: (650) 846-7635
                Fax: (650) 621-7817
                Email: pforeman@genencor.com
                Seq primer: LT-F1 primer.
                Location/Qualifiers
                1..815
                /organism="Hypocrea jecorina"
                /mol_type="mRNA"
                /strain="QMea"
                /db_xref="taxon:51453"
                /clone="trico25xf21"
                /dev_stage="mycelia"
                /clone_lib="T.reesei mycelial culture, Version 3 april"
                /notes="Vector: pREP3Y, Site_1: Not I/Sal I; Mycelial
                culture grown from 24 hrs to 6 days with varying Carbon
                and Nitrogen sources and concentrations."
ORIGIN
Query Match      68.9%; Score 18.6; DB 6; Length 815;
Best Local Similarity 84.0%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGCGCGCAATTGAAGTTATGTATC 25
    ||||| ||||| ||||| ||||| |||||
Db 138 GCGAGTGCATTCAGGTGATGTATC 114

RESULT 20
CF870869/c
LOCUS
DEFINITION      CF870869 815 bp mRNA linear EST 31-OCT-2003
                  trico25xf21.b1 T.reesei mycelial culture, Version 6 October 2003
                  Hypocrea jecorina cDNA clone trico25xf21, mRNA sequence.
ACCESSION      CF870869
VERSION        CF870869.1 GI:38125551
KEYWORDS       EST.
SOURCE         Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM       Hypocrea jecorina
                Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
                Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE      1 (bases 1 to 815)
AUTHORS        Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
                Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
                Ward, M. and Dean, R.A.
                Characterization of the protein processing and secretion pathways
                in a comprehensive set of expressed sequence tags from Trichoderma
                reesei
                FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
                Contact: Ralph A. Dean
                Fungal Genomics Laboratory
                North Carolina State University
                Campus Box 7251, Raleigh, NC 27695, USA

```

Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: LT-F1 primer.

FEATURES  
source  
1..815  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric025xf21"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN  
Query Match 68.9%; Score 18.6; DB 7; Length 815;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GC GGCCGAATTGCAAGTTATGTATC 25  
||| ||||||| |||||||  
Db 138 GCGACTGC AATTGAGTGATGATC 114

RESULT 21  
BH463620 818 bp DNA linear GSS 13-DEC-2001  
LOCUS BOHQ78TF BOHQ Brassica oleracea genomic clone BOHQ78, genomic survey sequence.  
DEFINITION BH463620 GI:17657947  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 818)  
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOHQ78TR  
COMMENT Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TP  
Class: sheared ends.  
Location/Qualifiers  
1..818  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHQ78"  
/clone\_lib="BOHQ"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 8; Length 818;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GC GGCCGAATTGAGTTATGTATCCT 27  
||| ||||||| |||||||

Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: LT-F1 primer.

FEATURES  
source  
1..815  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric025xf21"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN  
Query Match 68.9%; Score 18.6; DB 7; Length 815;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GC GGCCGAATTGCAAGTTATGTATC 25  
||| ||||||| |||||||  
Db 138 GCGACTGC AATTGAGTGATGATC 114

RESULT 21  
BH463620 818 bp DNA linear GSS 13-DEC-2001  
LOCUS BOHQ78TF BOHQ Brassica oleracea genomic clone BOHQ78, genomic survey sequence.  
DEFINITION BH463620 GI:17657947  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 818)  
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOHQ78TR  
COMMENT Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TP  
Class: sheared ends.  
Location/Qualifiers  
1..818  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHQ78"  
/clone\_lib="BOHQ"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 8; Length 818;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GC GGCCGAATTGAGTTATGTATCCT 27  
||| ||||||| |||||||

Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: LT-F1 primer.

FEATURES  
source  
1..815  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric025xf21"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN  
Query Match 68.9%; Score 18.6; DB 7; Length 815;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GC GGCCGAATTGCAAGTTATGTATC 25  
||| ||||||| |||||||  
Db 138 GCGACTGC AATTGAGTGATGATC 114

RESULT 21  
BH463620 818 bp DNA linear GSS 13-DEC-2001  
LOCUS BOHQ78TF BOHQ Brassica oleracea genomic clone BOHQ78, genomic survey sequence.  
DEFINITION BH463620 GI:17657947  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 818)  
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOHQ78TR  
COMMENT Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TP  
Class: sheared ends.  
Location/Qualifiers  
1..818  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHQ78"  
/clone\_lib="BOHQ"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 8; Length 818;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GC GGCCGAATTGAGTTATGTATCCT 27  
||| ||||||| |||||||

Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: LT-F1 primer.

FEATURES  
source  
1..815  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric025xf21"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN  
Query Match 68.9%; Score 18.6; DB 7; Length 815;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GC GGCCGAATTGCAAGTTATGTATC 25  
||| ||||||| |||||||  
Db 138 GCGACTGC AATTGAGTGATGATC 114

RESULT 21  
BH463620 818 bp DNA linear GSS 13-DEC-2001  
LOCUS BOHQ78TF BOHQ Brassica oleracea genomic clone BOHQ78, genomic survey sequence.  
DEFINITION BH463620 GI:17657947  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 818)  
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOHQ78TR  
COMMENT Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TP  
Class: sheared ends.  
Location/Qualifiers  
1..818  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHQ78"  
/clone\_lib="BOHQ"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN  
Query Match 68.9%; Score 18.6; DB 8; Length 818;  
Best Local Similarity 84.0%; Pred. No. 4.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GC GGCCGAATTGAGTTATGTATCCT 27  
||| ||||||| |||||||

Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: LT-F1 primer.

FEATURES  
source  
1..815  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric025xf21"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

FEATURES  
source

Location/Qualifiers  
1..919  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0718P03"  
/clone\_lib="ZM 0.7 1.5 KB"  
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 68.9%; Score 18.6; DB 9; Length 919;  
Best Local Similarity 84.0%; Pred. No. 4.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GGCGCGCAATTGAAGTTATGTATCT 27  
|||||  
Db 37 GGCCACCATATAGTTATGTATCT 61

## RESULT 24

CNS07G6H/c

LOCUS

DEFINITION  
Anopheles gambiae GSS SP6 end of clone 18509 of library NotreDamel  
from strain PEST of Anopheles gambiae (African malaria mosquito),  
genomic survey sequence.

ACCESSION

AL609371

VERSION

AL609371.1 GI:15915556

KEYWORDS

GSS.

SOURCE

ORGANISM

Anopheles gambiae (African malaria mosquito)

Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

Anopheles.

1 (bases 1 to 486)

Genoscope.

Direct Submission

Submitted (01-OCT-2001) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

2 (bases 1 to 486)

Roth, C.W., Brey, P.T., Ke, Z. and Collins, F.H.

Direct Submission

Submitted (01-OCT-2001) BBMI, Institut Pasteur, 25, rue du Dr.

Roux, Paris 75015, France

This clone is from an A. gambiae BAC library provided by F.H.

Collins and sequenced by Genoscope in collaboration with the

Laboratory of Biochem. and Biol. Molec. of Insects, Institut

Pasteur.

Location/Qualifiers

1..486

/organism="Anopheles gambiae"

/mol\_type="genomic DNA"

/strain="PEST"

/db\_xref="taxon:7165"

/clone="18E09"

/clone\_lib="NotreDamel"

/notes="end : SP6"

ORIGIN

Query Match 68.1%; Score 18.4; DB 9; Length 486;  
Best Local Similarity 95.0%; Pred. No. 5.5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CGCAATTGAAGTTATGTATC 25  
|||||

Db 472 CGCAATTGAAGTTATGCATC 453  
|||||

## RESULT 25

AZ338890

LOCUS

DEFINITION  
clone UUGC1M0070G08 F, genomic survey sequence.

ACCESSION

AZ338890

VERSION

AZ338890.1 GI:10412612

KEYWORDS

GSS.

SOURCE

ORGANISM

Mus musculus

(house mouse)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 628)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von

Niederhausern, A. and Wright, D. Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0070 Row: G Column: 08

Seq primer: CGTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 628.

Location/Qualifiers

1..628

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0070G08"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: FWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (GI|4732114|gb|AP129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

ORIGIN

Query Match 68.1%; Score 18.4; DB 8; Length 628;  
Best Local Similarity 95.0%; Pred. No. 5.7e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CAATTGAAGTTATGTATCT 27  
|||||

Db 128 CAATTGAAGTTATGTATCT 147  
|||||

```

RESULT 26
N26591/c
LOCUS
DEFINITION Yx91f05.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone
IMAGE:269121 3', mRNA sequence.
ACCESSION N26591
KEYWORDS N26591.1 GI:1140939
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 720)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisan,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 424
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: m3 -40 forward
High quality sequence stop: 424.
Location/Qualifiers
1. .720
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GB:3878763"
/db_xref="taxon:9606"
/clone="IMAGE:269121"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares melanocyte 2NbHM"
/notes="Vector: pT73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCACTCTGAAGTGGAGCGCGCGAGTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Patima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."

Query Match 68.1%; Score 18.4; DB 7; Length 720;
Best Local Similarity 95.0%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CAATTGAAGTTATGATCTCT 27
DB 302 CAATTGAGGTTATGATCTCT 283

RESULT 27
CR143861
LOCUS
DEFINITION Forward strand read from insert in 5'HPRT insertion targeting and
chromosome engineering clone MHPN246d22, genomic survey sequence.
ACCESSION CR143861
VERSION CR143861.1 GI:49891738
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 999)
ADAMS,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.
Direct Submission
Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICER
Location/Qualifiers
1. .999
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN246d22"
/clone_lib="MHPN"

Query Match 68.1%; Score 18.4; DB 9; Length 999;
Best Local Similarity 95.0%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CAATTGAAGTTATGATCTCT 27
DB 596 CAATTGAACTTATGATCTCT 615

RESULT 28
BU112231/c
LOCUS
DEFINITION BU112231 768 bp mRNA linear EST 25-NOV-2002
603002410F1 CSEQCHL14 Gallus gallus cDNA clone Chest10c4 5', mRNA
sequence.
ACCESSION BU112231
VERSION BU112231.1 GI:25316259
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 768)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curt. Biol. 12 (22), 1965-1969 (2002)
22335534
MEDLINE
12445392
PUBMED
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .768
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="Chest10c4"
/dev_stage="22"
/lab_host="DH10B"
/clone_lib="CSEQCHL14"
/notes="Organ: head; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
[Stratagene] vector to accommodate cDNA produced with the
T-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent #
6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI."

```

Ligate in double stranded adaptor containing BsgI and BamHI sites (5'ggcgcgtgcagcccgatccgaaaaag][5'aattcttttttgcgatccgggtgcgc]

## ORIGIN

Query Match 67.4%; Score 18.2; DB 5; Length 768;  
 Best Local Similarity 87.0%; Pred. No. 7.3e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GCCGCAATTGAAGTTATGTATCC 26  
 ||||| ||||| ||||| ||||| |||||  
 Db 196 GCTGCAATGGAAGTTACGTATCC 174

RESULT 29  
 BU335623/c 788 bp mRNA linear EST 28-NOV-2002  
 LOCUS 603496949F1 CSBQCHN65 Gallus gallus cDNA clone CHEST408a13 5', mRNA  
 DEFINITION sequence.  
 ACCESSION BU335623  
 VERSION BU335623.1 GI:25843624  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 REFERENCE 1 (bases 1 to 788)  
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 TITLE A Comprehensive Collection of Chicken CDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

## FEATURES

source  
 1. .788  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST408a13"  
 /tissue\_type="whole embryo"  
 /dev\_stage="10"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN65"  
 /note="Organ: whole embryo; Vector: pBluescript II KS(+);  
 Site 1: EcoRI; Site 2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

## ORIGIN

Query Match 67.4%; Score 18.2; DB 5; Length 788;  
 Best Local Similarity 87.0%; Pred. No. 7.3e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  
 Db

RESULT 30  
 BU381772/c  
 LOCUS 603496949F1 CSBQCHN75 Gallus gallus cDNA clone CHEST859c24 5', mRNA  
 DEFINITION sequence.  
 ACCESSION BU381772  
 VERSION BU381772.1 GI:25889773  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 REFERENCE 1 (bases 1 to 825)  
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 TITLE A Comprehensive Collection of Chicken CDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
 source

1. .825  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST859c24"  
 /dev\_stage="16"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN75"  
 /note="Organ: trunks; Vector: pBluescript II KS(+);  
 Site 1: EcoRI; Site 2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was blunted, ligated to NotI adapters, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

## ORIGIN

Query Match 67.4%; Score 18.2; DB 5; Length 825;  
 Best Local Similarity 87.0%; Pred. No. 7.4e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GCCGCAATTGAAGTTATGTATCC 26  
 ||||| ||||| ||||| ||||| |||||  
 Db 199 GCTGCAATGGAAGTTACGTATCC 177

## RESULT 31

CNS02N5T  
 LOCUS 838 bp DNA linear GSS 01-SBP-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence P7 end of clone  
 150J10 of library G from Tetraodon nigroviridis, genomic survey



```

sequence.
AL204986      1  GI:7863805
GSS; genome survey sequence.
Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
REFERENCE
AUTHORS      Roest Crolius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C.,
              Bernot.A., Pizames.C., Wincker.P., Brottier.P., Quetier.F.,
              Saurin.W. and Weissenbach.J.
TITLE        Estimate of human gene number provided by genome-wide analysis
              using Tetraodon nigroviridis DNA sequence
JOURNAL      Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE      20296633
PUBMED       10835645
REFERENCE
AUTHORS      2
              Roest Crolius.H., Jaillon.O., Dasilva.C., Ozouf-Costaz,C.,
              Pizames.C., Fischer.C., Bouneau.L., Billault.A., Quetier.F.,
              Saurin.W., Bernot.A. and Weissenbach.J.
TITLE        Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetraodon nigroviridis
JOURNAL      Genome Res. 10 (7), 939-949 (2000)
MEDLINE      20359837
PUBMED       10899143
REFERENCE
AUTHORS      3 (bases 1 to 838)
              Genoscope.
DIRECT SUBMISSION
              Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT      This sequence is a single read and was generated as part of a large
              scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1. .838
   /organism="Tetraodon nigroviridis"
   /mol_type="genomic DNA"
   /db_xref="taxon:99883"
   /clone="150J10"
   /clone_lib="G"
   /notes="Genoscope sequence ID : COAGI50DE05LP1-end : T7"
ORIGIN
Query Match      67.4%; Score 18.2; DB 9; Length 838;
Best Local Similarity 74.1%; Pred. No. 7.4e+02;
Matches 20; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGCCGCAATTCGAAGTTATGTATCCT 27
Db 773 GCTGTCACAACTGAAGTWTGATCAT 799

RESULT 32
LOCUS CL023772      1329 bp      DNA      linear      GSS 31-DEC-2003
DEFINITION CH216-1718 RM1.1 CH216 Xenopus tropicalis genomic clone CH216-1718,
genomic survey sequence.
ACCESSION CL023772
VERSION CL023772.1 GI:40465585
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 1329)
REFERENCE
AUTHORS      Kremtzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
              Mardis,E. and Wilson.R.

A physical map of the xenopus tropicalis genome
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM1 TACGACTCATTATGGGAGA
Class: BAC ends
High quality sequence start: 131
High quality sequence stop: 271.
FEATURES
source
1..1329
   Location/Qualifiers
   /organism="Xenopus tropicalis"
   /mol_type="genomic DNA"
   /strain="Nigerian frog"
   /db_xref="taxon:8364"
   /clone="CH216-1718"
   /sex="male"
   /cell_line="Stock 248 F7A2, inbred N7"
   /clone_lib="CH216"
   /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
   BAC library"
ORIGIN
Query Match      67.4%; Score 18.2; DB 9; Length 1329;
Best Local Similarity 87.0%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CCGCAATTCGAAGTTATGTATCCT 27
Db 190 CCGAAATTCGATTATGTATCCT 212

RESULT 33
LOCUS BY592030/c
DEFINITION BY592030 RIKEN full-length enriched, adult inner ear Mus musculus
CDNA clone F930021E21 3', mRNA sequence.
ACCESSION BY592030
VERSION BY592030.1 GI:26927212
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 205)
REFERENCE
AUTHORS      Okazaki,Y., Furuno.M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
              Nikaido,I., Osato.N., Saito,R., Suzuki,H., Yamanaka,I.,
              Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
              Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
              Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
              Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
              Chothia,C., Corbani,L.E., Cousins,S., Dalla.E., Dragani,T.A.,
              Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
              Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
              Gustinch,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
              Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
              Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
              Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
              Numata,K., Okido,T., Pavan,W.J., Perte,G., Pesole,G.,
              Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
              Ravasi,T., Reed,J.C., Reid,J., Ring,B.Z., Ringwald,M.,
              Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K.,
              Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
              Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
              Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
              Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
              Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
              Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
              Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Inotani,K., Ishii,Y.,
              Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
              Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,

```

**TITLE** Rogers, J., Birney, E. and Hayashizaki, Y.  
**JOURNAL** Analysis of the mouse transcriptome based on functional annotation  
**MEDLINE** of 60,770 full-length cDNAs  
**PUBMED** Nature 420, 563-573 (2002)  
**COMMENT** 22354683  
 12466851  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9226  
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
 Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
 Ono, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
 Hayashizaki, Y. Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Kirk W. Beisel ( Boys Town National  
 Research Hospital 555 North 30th Street Omaha, NE 68131 USA ) whose  
 assistance we gratefully acknowledge.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.

# FEATURES

source  
 1. .205  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="F930021E21"  
 /tissue\_type="inner ear"  
 /dev\_stage="adult"  
 /clone\_lib="RIKEN full-length enriched, adult inner ear"

# ORIGIN

Query Match 66.7%; Score 18; DB 5; Length 205;  
 Best Local Similarity 80.8%; Pred. No. 7.9e+02;  
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CGGCCGCAATTGAAGTTATGTATCCT 27  
 |||||  
 Db 50 CGGCTGAATTGACCTTATCTATCCT 25

RESULT 34  
 CB676984  
 LOCUS CB676984 508 bp mRNA linear EST 09-APR-2003  
 DEFINITION OSJNEE13120.f OSJNEE Oryza sativa (japonica cultivar-group) cDNA  
 clone OSJNEE13120 5', mRNA sequence.  
 ACCESSION CB676984  
 VERSION CB676984.1 GI:29680709  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.  
 1 (bases 1 to 508)  
 Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,  
 Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea  
 Unpublished (2003)  
 Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 13 row: 1 column: 20  
 Seq primer: gta aaa cga cgg cca gtc.

# FEATURES

source  
 1. .508  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNEE13120"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEE"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
 XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

# ORIGIN

Query Match 66.7%; Score 18; DB 6; Length 508;  
 Best Local Similarity 80.8%; Pred. No. 8.7e+02;  
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCGCGCGCAATTGAAGTTATGTATCCT 26  
 |||||  
 Db 256 GCAGCGCATTGCGAGGTGTATCCT 281

RESULT 35  
 AQ654959  
 LOCUS AQ654959 561 bp DNA linear GSS 22-JUN-1999  
 DEFINITION Sheared DNA-21K22.TR Sheared DNA Trypanosoma brucei genomic clone  
 Sheared DNA-21K22, genomic survey sequence.  
 ACCESSION AQ654959  
 VERSION AQ654959.1 GI:5148145  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.

REFERENCE  
 AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,  
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S.,  
 Donelson, J., Fraser, C. and Adams, M.  
 Determination of clone end sequences from Trypanosoma brucei GUTat  
 10.1 sheared DNA library  
 Unpublished (1999)  
 JOURNAL Other GSSs: Sheared DNA-21K22.TF  
 COMMENT Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared

DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/cdb/mbd/tbdb/>.  
Seq primer: M13-Reverse  
Class: shotgun.

#### FEATURES

Location/Qualifiers  
1..561  
/organism="Trypanosoma brucei"  
/mol\_type="genomic DNA"  
/strain="TREU927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone="Sheared DNA-21K22"  
/clone\_lib="Sheared DNA"  
/note="Vector: pUC18; Site 1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

#### ORIGIN

Query Match 66.7%; Score 18; DB 8; Length 561;  
Best Local Similarity 80.8%; Pred. No. 8.8e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGCGCCGAATTGACGTATGATCTCT 27  
|||||  
DB 339 CGCGCCGAATTGATGATCTCT 364

#### RESULT 36

BQ135922  
LOCUS  
DEFINITION  
BQ135922 562 bp mRNA linear EST 23-APR-2002  
clone NF023D01EC1F1012 Elicited cell culture Medicago truncatula cDNA

ACCESSION  
BQ135922  
VERSION  
BQ135922.1 GI:20272046  
KEYWORDS  
EST.

#### SOURCE

Medicago truncatula (barrel medic)  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

#### REFERENCE

1 (bases 1 to 562)  
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research  
Unpublished (2000)

#### JOURNAL

Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7302  
Fax: 580 221 7380

Email: [radixon@noble.org](mailto:radixon@noble.org)  
Insert Length: 562 Std Error: 0.00  
Plate: 023 row: D column: 01  
Seq primer: TCACACAGGAACAGCTATCAC.

#### FEATURES

Location/Qualifiers  
1..562  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF023D01EC"  
/tissue\_type="Cell cultures derived from root tissues"  
/dev\_stage="Cell suspensions were subcultured every 14

days. Cells were induced six days after subculture"  
/clone\_lib="Elicited cell culture"  
/note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5, 1, 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

#### ORIGIN

Query Match 66.7%; Score 18; DB 5; Length 562;  
Best Local Similarity 80.8%; Pred. No. 8.8e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGCGCCGAATTGACGTATGATCTCT 27  
|||||  
DB 103 CAGCGCAATTGATGATCTCT 128

#### RESULT 37

AQ644499  
LOCUS  
DEFINITION  
AQ644499 599 bp DNA linear GSS 08-JUL-1999  
RPC193-DpnII-28M12.TV RPC193-DpnII Trypanosoma brucei genomic clone  
RPC193-DpnII-28M12, genomic survey sequence.

ACCESSION  
AQ644499  
VERSION  
AQ644499.1 GI:5121209  
KEYWORDS  
GSS.

#### SOURCE

Trypanosoma brucei  
Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

#### REFERENCE

1 (bases 1 to 599)  
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J., Fraser, C. and Adams, M.

Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93 Library for gene discovery and sequence-ready map construction  
Unpublished (1999)

#### JOURNAL

Other GSSs: RPC193-DpnII-28M12.TJ  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: [nelsayed@tigr.org](mailto:nelsayed@tigr.org)  
Clones and high density filters may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu>). BAC end sequences search page: <http://www.tigr.org/cdb/mbd/tbdb/>.

Seq primer: T7  
Class: BAC ends.

Location/Qualifiers  
1..599

#### FEATURES

##### source

/organism="Trypanosoma brucei"  
/mol\_type="genomic DNA"  
/strain="TREU927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone\_lib="RPC193-DpnII-28M12"  
/clone="RPC193-DpnII-28M12"  
/note="Vector: pBACE3.6; Site 1: Bam HI; Site 2: Bam HI; Constructed for The Institute for Genomic Research by Bohui Zhao in Pieter de Jong's laboratory (Roswell Park Cancer Institute, Buffalo, NY). Briefly, Trypanosoma brucei TREU927/4 GUTat 10.1 agarose embedded DNA was partially digested with a combination of Eco RI and Eco RI methylase (RPC193-EcoRI segment) or Dpn II (RPC193-DpnII segment). High molecular weight fragments were ligated in pBACE3.6 vector digested with Eco RI or Bam HI, respectively. The average insert size is 141 kb. Total coverage (both segments): > 90 X the haploid non-minichromosomal genome."

#### ORIGIN

Query Match 66.7%; Score 18; DB 8; Length 599;

```

Best Local Similarity 80.8%; Pred. No. 8.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGGCCGCAATTGAAGTTATGATCTCT 27
    |||||
Db 540 CGGCCGCAATTGATGATCTCT 565

RESULT 38
AQ787083
LOCUS
DEFINITION
HS_5480_A1_B02_T7A_RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1056 Col=3 Row=C, genomic survey sequence.
ACCESSION
AQ787083
VERSION
AQ787083.1 GI:5694707
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 621)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL
MEDLINE
99380589
PUBMED
10449764
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1056 row: C column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 621.
FEATURES
source
1..621
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=1056 Col=3 Row=C"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

ORIGIN
Query Match 66.7%; Score 18; DB 8; Length 621;
Best Local Similarity 80.8%; Pred. No. 8.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CGGCCGCAATTGAAGTTATGATCTCT 26
    |||||
Db 332 GCTGCACACCTGAAGTTATGTTTCC 357

RESULT 39
BF649189
LOCUS
DEFINITION
OSJNEB16G03.f OSJNEB Oryza sativa (Japonica cultivar-group) cDNA
clone OSJNB16G03 5', mRNA sequence.
ACCESSION
CB651428
VERSION
CB651428.1 GI:29646421
KEYWORDS
EST.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Zrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 719)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
JOURNAL
COMMENT
Contact: Rod Wing

```

```

DEFINITION
NF055A06ECIF1039 Elicited cell culture Medicago truncatula cDNA
clone NF055A06EC 5', mRNA sequence.
ACCESSION
BF649189
VERSION
BF649189.1 GI:11914319
KEYWORDS
EST.
SOURCE
Medicago truncatula (barrel medic)
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 625)
Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
Center for Medicago Genomics Research
Unpublished (2000)
JOURNAL
COMMENT
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 625 Std Error: 0.00
Plate: 055 row: A column: 06
Seq primer: TCACACGAGAAACAGCTATGAC.
FEATURES
source
1..625
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
days. Cells were induced six days after subculture"
/clone_lib="Elicited cell culture"
/notes="Vector: Lambda Zap; Cells were induced with yeast
cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."

ORIGIN
Query Match 66.7%; Score 18; DB 2; Length 625;
Best Local Similarity 80.8%; Pred. No. 8.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGGCCGCAATTGAAGTTATGATCTCT 27
    |||||
Db 409 CAGCGCAAAATGAATTTATGATCTCT 434

RESULT 40
CB651428
LOCUS
DEFINITION
OSJNEB16G03.f OSJNEB Oryza sativa (Japonica cultivar-group) cDNA
clone OSJNB16G03 5', mRNA sequence.
ACCESSION
CB651428
VERSION
CB651428.1 GI:29646421
KEYWORDS
EST.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Zrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 719)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
JOURNAL
COMMENT
Contact: Rod Wing

```

Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtcg  
BACKWARD: gga aac agc tat gac cat g  
Plate: 16 row: G column: 03  
Seq primer: gta aaa cga cgg cca gtcg.

# FEATURES

Location/Qualifiers  
1..719  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNEb16G03"  
/tissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSJNEb"  
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
XhoI; 24 hrs after inoculation with Rice Blast (Che  
86061)"

# ORIGIN

Query Match 66.7%; Score 18; DB 6; Length 719;  
Best Local Similarity 80.8%; Pred. No. 9.1e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 GCGGCGCAATTGAGCTTATGTATCC 26  
||| ||||| ||||| ||||| |||||  
Db 246 GCAGCGCATTCGAGGTGTATCC 271

Search completed: January 7, 2005, 12:37:39  
Job time : 1617.95 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 228.421 Seconds  
(without alignments)  
689.440 Million cell updates/sec

Title: US-09-786-502A-8  
Perfect score: 30  
Sequence: 1 tcgagagatcttcaggagcgataggctgc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	AA10271	Human CD2
2	30	100.0	30	ABA92025	Human CD2
3	20.4	68.0	454	ADK98583	Human lmm
4	20.4	68.0	875	AAT27650	CD2-CD28
5	20.4	68.0	1064	AAX90349	Human CD2
6	20.4	68.0	1514	AAQ21167	Human CD2
7	20.4	68.0	1514	AAT36257	Human CD2
8	20.4	68.0	1514	AAT14707	Human CD2
9	20.4	68.0	1514	AAV63445	Human CD2
10	20.4	68.0	1514	AAV81203	Human CD2
11	20.4	68.0	1514	AAAS0581	Human cel
12	20.4	68.0	1514	AAX29325	Human CD2
13	20.4	68.0	1514	AAAS03175	Human lym
14	20.4	68.0	1514	ABA99035	Human CD2
15	20.4	68.0	1514	ADD25537	Binding d
16	20.4	68.0	1514	AD131990	Human cDN
17	20.4	68.0	1514	ADO49343	Human CD2
18	20.4	68.0	1574	AAN90607	CD28 anti
19	20.4	68.0	2405	AAT27652	LFA-1-bet
20	20.4	68.0	3585	AAT27651	LFA-1-alp
21	20.4	68.0	3803	AAL49546	Human CD2

C 22	20.4	68.0	3804	8	ABV75351	Human CD2
C 23	20.4	68.0	3804	9	ADA02991	Human CD2
C 24	20.4	68.0	3804	10	ADB72729	Human CD2
C 25	20.4	68.0	3804	10	ADC85471	Human CD2
C 26	20.4	68.0	3804	12	ADM74586	Human car
C 27	20.4	68.0	3804	12	ADN05855	Antipeori
C 28	20.4	68.0	3806	6	AAD21976	Human CD2
C 29	20.4	68.0	51365	9	ADA02990	Human CD2
C 30	20.4	68.0	51365	10	ADB72728	Human CD2
C 31	20.4	68.0	51365	10	ADC85470	Human CD2
C 32	20.4	68.0	51365	12	ADM74585	Human car
C 33	20.2	67.3	36	12	ADL67233	Human CD2
C 34	20	66.7	31	6	ABN88948	Human CD2
C 35	20	66.7	31	10	ACF80096	Human CD8
C 36	20	66.7	31	10	ACC70422	PCR prime
C 37	19.4	64.7	30	2	AAT90526	Human CD2
C 38	19.4	64.7	64	2	AAT90529	T cell re
C 39	19.4	64.7	852	8	ADA69416	Rice gene
C 40	19.4	64.7	1156	5	AA881566	DNA encod
C 41	19.4	64.7	2217	5	AA885808	DNA encod
C 42	19.4	64.7	2217	5	AA883371	DNA encod
C 43	19.4	64.7	3544	5	AA881565	DNA encod
C 44	19.4	64.7	6335	6	ABK83722	Human CDN
C 45	19.4	64.7	6335	11	ADN95256	Human BEC

ALIGNMENTS

RESULT 1  
AA10271  
ID AA10271 standard; DNA; 30 BP.  
XX  
AC AA10271;  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE Human CD28 fragment downstream PCR primer.

XX  
OS Homo sapiens.  
XX  
PN WO200014257-A1.  
XX  
PD 16-MAR-2000.  
XX  
PF 03-SEP-1999; 99WO-US020349.  
XX  
PR 04-SEP-1998; 98US-0099138P.  
XX  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
PI Sadelain M, Bander NH, Gong M;  
XX  
DR WPI; 2000-257002/22.

XX  
PT A fusion receptor composition having the structure:prostate-specific membrane antigen-single chain variable fragment:optional connector:cytoplasmic domain, useful for treatment of cancer.

XX  
PS Example 6; Page 15; 25pp; English.

XX  
CC The invention relates to a novel fusion receptor composition having the structure: PSMA (prostate-specific membrane antigen)-scfv (single chain variable fragment):optional connector:cytoplasmic domain, where the fusion receptor is effective when expressed in a T-cell to promote a cellular immune response to PSMA. The PSMA-scfv is a single-chain antibody cloned from the V region genes of a hybridoma specific for PSMA, such as J591. The optional connector is provided to give a spacing

CC between the PSMA-scfv and the cytoplasmic domain, such that both retain  
 CC substantial function. The cytoplasmic domain directs the function of the  
 CC fusion receptor and is generally the cytoplasmic domain of a molecule  
 CC which functions as a transducer of a mammalian immune response in the  
 CC presence of an MHC (major histocompatibility complex)-peptide complex or  
 CC costimulatory factor. Examples of cytoplasmic domains that may be  
 CC employed in the present invention include the T-cell receptor gamma-chain  
 CC cytoplasmic domain and the CD28 cytoplasmic domain. In a method of the  
 CC invention, an expression vector encoding the fusion receptor is  
 CC transduced into primary T-lymphocytes obtained from the patient to be  
 CC treated. The transduced lymphocytes are returned to the patient where  
 CC they secrete interleukin-2 (IL-2) and proliferate in response to PSMA-  
 CC positive cells. The resulting cytotoxic lymphocytes specifically lyse  
 CC cells expressing PSMA and can thus be used to target PSMA-positive tumour  
 CC cells. The fusion receptor promotes a cellular immune response to PSMA  
 CC and is useful for the treatment of prostate cancer and other cancers that  
 CC express PSMA. Sequences AAA10270-A10271 represent PCR primers used in an  
 CC exemplification of the present invention to amplify cDNA encoding the  
 CC cytoplasmic and transmembrane domains and part of the extracellular  
 CC domain of human CD28. This was used to construct a gene encoding a fusion  
 CC receptor comprising the PSMA-scfv and the CD28 fragment  
 XX  
 SQ Sequence 30 BP; 6 A; 6 C; 11 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 3; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 0.00094;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
 DB 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30

## RESULT 2

ABA92025  
 ID ABA92025 standard; DNA; 30 BP.  
 XX  
 AC ABA92025;  
 XX  
 DT 23-MAY-2002 (first entry)  
 XX  
 DE Human CD28 cDNA downstream PCR primer.  
 XX  
 KW CD28; T cell; receptor; human; antitumour; immunostimulant; cancer;  
 KW therapy; neuroblastoma; melanoma; sarcoma; small lung carcinoma;  
 KW brain tumour; disialoganglioside GD2; imaging; PCR; primer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002018783-A1.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 30-SEP-1997; 97US-00940544.  
 XX  
 PR 20-MAR-1997; 97WO-US004427.  
 XX  
 PA (SADE/) SADELAIN M.  
 PA (CHEU/) CHEUNG N V.  
 PA (KRAU/) KRAUSE A.  
 PA (GUOH/) GUO H.  
 XX  
 PI Sadelain M, Cheung NV, Krause A, Guo H;  
 XX  
 DR WPI; 2002-239251/29.  
 XX  
 XX New fusion protein, useful for inducing host immune response, comprises  
 PT variable region of light chain of an antibody linked to variable region  
 PT of antibody, CD28 receptor signaling domain and transmembrane domain.  
 XX  
 XX Example 3; Page 4; 9pp; English.  
 PS  
 PS The present sequence is that of a downstream primer, used with the

CC upstream primer given in ABA92024, in the PCR amplification of a segment  
 CC of human T cell surface receptor CD28 cDNA. The amplified segment encodes  
 CC part of the extracellular domain of CD28, and the transmembrane and the  
 CC cytoplasmic domains. Plasmid pbsCD28 was used as template. The 5' primer  
 CC contains an NotI site, while the 3' primer contains a BamHI site, and  
 CC these sites in the PCR product were used to facilitate insertion into  
 CC retroviral vector SFG. A cell-surface molecule capable of CD28 signalling  
 CC in T cells interacting with disialoganglioside GD2 tumours was  
 CC constructed. This comprised the antigen-binding site of a GD2-specific  
 CC antibody and the transmembrane and signalling domains of the CD28  
 CC molecule. This is an example of fusion proteins of the invention. T cells  
 CC comprising a single chain Fv antibody linked to CD28 receptor. T cells  
 CC expressing such fusion proteins exhibit enhanced survival when  
 CC reintroduced to an in vivo environment. They can be used to induce an  
 CC immune response to cells, particularly tumour cells which express the  
 CC antigen to which the antibody is specific. Cells expressing a fusion  
 CC protein including an anti-GD2 scFv are useful for treatment of melanomas,  
 CC neuroblastomas, small lung carcinoma, sarcomas and brain tumours that  
 CC express GD2 as a surface antigen. Cells expressing the fusion proteins of  
 CC the invention can also be used for in vitro purging of stem cells or bone  
 CC marrow and for in vivo targeting of tumour cells and other antigen-  
 CC bearing cells for imaging  
 XX  
 SQ Sequence 30 BP; 6 A; 6 C; 11 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 6; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 0.00094;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
 DB 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30

## RESULT 3

ADK98583/C  
 ID ADK98583 standard; cDNA; 454 BP.  
 XX  
 AC ADK98583;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Human immune response associated protein IRAP-11 cDNA.  
 XX  
 KW immune response associated protein; IRAP; antiarteriosclerotic;  
 KW cytosolic; neuroprotective; antiparkinsonian; hepatotropic;  
 KW cerebroprotective; antiinflammatory; nootropic; vasotropic;  
 KW arteriosclerosis; cirrhosis; cancer; stroke; Alzheimer's disease;  
 KW Parkinson's; Crohn's; gene therapy; human; ss; gene.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004020593-A2.  
 XX  
 PD 11-MAR-2004.  
 XX  
 PF 26-AUG-2003; 2003WO-US026988.  
 XX  
 PR 30-AUG-2002; 2002US-0407561P.  
 PR 11-SEP-2002; 2002US-0410178P.  
 PR 13-SEP-2002; 2002US-0410571P.  
 PR 18-OCT-2002; 2002US-0419906P.  
 PR 25-OCT-2002; 2002US-0421445P.  
 XX  
 PA (INCY-) INCYTE CORP.  
 XX  
 PI Ramkumar J, Swarnakar A, Elliott VS, Hafalia AJA, Richardson TW;  
 PI Lee SY, Lindquist EA, Marquis JP, Chawla NK, Khare R, Becha SD;  
 XX  
 DR WPI; 2004-239178/22.  
 DR N-PSDB; ADK98548.  
 XX  
 PT New isolated immune response associated proteins (IRAP) polypeptide and



PT polynucleotide, useful for diagnosing and/or treating disorders with  
 PT aberrant expression of IRAP, such as arteriosclerosis, cirrhosis, cancer  
 XX and stroke.

PS Claim 5; SEQ ID NO 46; 207bp; English.

XX The invention relates to a novel isolated immune response associated  
 CC protein (IRAP) comprising any of 35 fully defined sequences given in the  
 CC specification. The polypeptide of the invention demonstrates  
 CC antiarteriosclerotic, cytostatic, neuroprotective, antiparkinsonian,  
 CC hepatotropic, cerebroprotective, antiinflammatory, neurotropic and  
 CC vasotropic activities and may be useful for treating a disease or  
 CC condition associated with decreased expression or overexpression of  
 CC functional immune response associated proteins, while the antibody is  
 CC useful for diagnosing a condition or disease associated with the  
 CC expression of IRAP, such as arteriosclerosis, cirrhosis, cancer, stroke,  
 CC Alzheimer's disease, Parkinson's disease and Crohn's disease.  
 CC Furthermore, the molecules of the invention may be utilised during gene  
 CC therapy procedures. The current sequence is that of a human IRAP cDNA of  
 CC the invention.

XX Sequence 454 BP; 109 A; 142 C; 105 G; 98 T; 0 U; 0 Other;

Query Match 68.0%; Score 20.4; DB 12; Length 454;  
 Best Local Similarity 95.5%; Pred. No. 31;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30  
 Db 391 CGTGTCTCAGGAGCGATAGGCTGC 370

RESULT 4  
 AAT27650/C  
 ID AAT27650 standard; cDNA; 875 BP.

AC AAT27650;

DT 24-AUG-1996 (first entry)

DE CD2-CD28 chimeric receptor.

XX Human; chimeric receptor; CD2; extracellular domain; CD28;  
 KW transmembrane region; intracellular domain; retrovirus; vector;  
 KW packaging cell culture; gene transfer; LFA-3; selective activation;  
 KW cytotoxic T-lymphocyte; antitumour; immunotherapy; cancer; therapy;  
 KW virucide; ss.

OS Homo sapiens.

XX WO9613584-A1.

PN 09-MAY-1996.

XX 01-NOV-1995; 95WO-US014171.

XX 01-NOV-1994; 94US-00332993.

PR 07-JUN-1995; 95US-00480577.

XX (TARG-) TARGETED GENETICS CORP.

XX Feldhaus AL, Jones LA;

PI WPI; 1996-239497/24.

XX TH independent cytotoxic T lymphocytes useful in tumour or viral therapy  
 PT - are activated by adhesion to antigen displaying cells.

XX Example 1; Page 54-55; 82pp; English.

XX The sequence is a BamHI-SalI fragment encoding a chimeric receptor,  
 CC composed of the extracellular region (XR) of CD2, fused via the  
 CC transmembrane region (TM) of CD28 to the intracellular region (IC28) of

CC CD28. TM and IC28 coding sequences of human CD28 (bp 556-767) are  
 CC amplified by PCR from human CD28 cDNA, and the XR of CD2 (bp 1-627) is  
 CC amplified by PCR from human CD2 cDNA. The sequences are cloned to form  
 CC plasmid pSKcytoCD28, and subcloned in HyTK-CMV-cat to form HyTK-CMV-  
 CC CD2/CD28. This retrovirus vector is introduced into a psi-2 ecotropic  
 CC packaging cell culture to give a stable PA317 cell line producing high  
 CC titers of infectious retrovirus particles, which are then used to infect  
 CC T-lymphocytes. The resulting cytotoxic T-lymphocytes (CTLs) are  
 CC selectively activatable, and have a lessened dependence on helper T-  
 CC lymphocytes (TH cells) and/or growth factors. The chimeric receptor binds  
 CC to the CD2 ligand (leukocyte function-associated antigen-3 or LFA-3), but  
 CC IC28 transmits the proliferative signal of CD28. The CTLs mediate lysis  
 CC of a target cell bearing a cognate antigen, and are useful in directed  
 CC cancer immunotherapy, and in treatment of viral infections

XX Sequence 875 BP; 277 A; 188 C; 202 G; 208 T; 0 U; 0 Other;

Query Match 68.0%; Score 20.4; DB 2; Length 875;  
 Best Local Similarity 95.5%; Pred. No. 35;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30  
 Db 868 CGTGTCTCAGGAGCGATAGGCTGC 847

RESULT 5  
 AAX90349/C  
 ID AAX90349 standard; DNA; 1064 BP.

AC AAX90349;

DT 24-SEP-1999 (first entry)

DE Human CD28 gene DNA sequence #2.

XX CD28; inhibition; antisense oligonucleotide; interleukin 2; IL-2;  
 KW immune system mediated disease; gamma-interferon; IL-8; ss.

OS Homo sapiens.

PN US5932556-A.

XX 03-AUG-1999.

XX 18-SEP-1995; 95US-00529878.

XX 09-FEB-1995; 95US-00387041.

PR 18-SEP-1995; 95US-00529878.

XX (TAMR/) TAM R C.

XX Tam RC;

XX WPI; 1996-384228/38.

DR P-PSDB; AAY24470.

XX Oligo:nucleotide which reduces CD28 gene expression in T cells - for  
 PT treating immune system diseases, e.g. graft vs. host disease, septic  
 PT shock, psoriasis, etc.

XX Disclosure; Fig 1C; 45pp; English.

XX The present invention describes a method for inhibiting the expression of  
 CC CD28, IL-2, gamma-interferon or IL-8 in a mammal. The method comprises  
 CC subcutaneous administration of an oligonucleotide (OGN). AAX90288 to  
 CC AAX90291 represent specifically claimed OGNs for use in the method. The  
 CC OGNs are used for the treatment of immune system-mediated diseases. The  
 CC present sequence represents a DNA sequence of the human CD28 gene from  
 CC the present invention

XX Sequence 1064 BP; 289 A; 255 C; 230 G; 290 T; 0 U; 0 Other;

Query Match	68.0%;	Score 20.4;	DB 2;	Length 1064;
Best Local Similarity	95.5%;	Pred. No. 36;		
Matches	21;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	9	CTTGTCAGGAGCGATAGGCTGC	30	
Db	316	CGTGTCAGGAGCGATAGGCTGC	295	
RESULT 6				
AAQ21167/c				
ID	AAQ21167	standard; DNA; 1514 BP.		
XX	AAQ21167;			
XX				
XX	25-MAR-2003	(revised)		
DT	21-MAY-1992	(first entry)		
XX				
XX	Human CD28	antigen coding sequence.		
DE				
XX				
XX	cloning technique;	cell surface antigen; immunodiagnosis; tumour; ss.		
KW				
XX				
OS	Homo sapiens.			
XX				
XX	Key	Location/Qualifiers		
FH	CDS	100..762		
FT		/*tag= a		
FT		/product= "CD28"		
FT	sig_peptide	100..153		
FT		/*tag= b		
XX				
XX	WO9201049-A.			
PN				
XX				
XX	23-JAN-1992.			
PD				
XX				
XX	13-JUL-1990;	90US-00553759.		
PF				
XX				
PR	13-JUL-1990;	90US-00553759.		
XX				
XX	(GEO )	GEN HOSPITAL CORP.		
PA				
XX				
XX	Seed B, Aruffo A, Amiot M;			
PI				
XX				
XX				
DR	WPI; 1992-056864/07.			
DR	P-PSDB; AAR20805.			
XX				
XX	New CD53 cell surface antigen and DNA encoding it - for immuno-therapy			
PT	and diagnosis of haematopoietic neoplasms, etc.			
XX				
PS	Example 3; Fig 7; 160pp; English.			
XX				
CC	A library of recombinants having inserts greater than 0.8kb in size was			
CC	prepared in the pBl3M vector (see AAQ21166) from 1 microgram of polyA			
CC	(plus) RNA isolated from the human lymphoblastoid cell line JY. The			
CC	library was screened for CD28 cDNA clones which were isolated using the			
CC	antibody enrichment method (see e.g. AAQ21164 or AAQ21165). After the			
CC	third transfection, COS cells were panned with a specific anti-CD28 Ab. A			
CC	third supernatant was generated and transformed into E.coli. DNA was			
CC	prepared from the resulting colonies and transfected into COS cells.			
CC	Surface expression of CD28 antigen was detected in 3 of the 8 transfected			
CC	cultures by indirect immunofluorescence. The cDNA insert from one of the			
CC	positive clones was sequenced. (Updated on 25-MAR-2003 to correct PA			
CC	field.)			
XX				
XX				
SQ	Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;			
Query Match	68.0%;	Score 20.4;	DB 2;	Length 1514;
Best Local Similarity	95.5%;	Pred. No. 37;		
Matches	21;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	9	CTTGTCAGGAGCGATAGGCTGC	30	
Db	766	CGTGTCAGGAGCGATAGGCTGC	745	

```
XX AC AAT14707;
XX DT 25-MAR-2003 (revised)
XX DT 30-OCT-1996 (first entry)
XX DE Human CD28 cDNA.
XX KW Cell surface antigen; cloning; immunoselection; immunotherapy; therapy;
XX KW diagnosis; vector; pIH3M; CD28; COS; T-lymphocyte; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 100..762
XX FT sig_peptide 100..153
XX FT mat_peptide 154..759
XX FT /*tag= c
XX PN US5506126-A.
XX PD 09-APR-1996.
XX PF 18-OCT-1993; 93US-00139273.
XX PR 25-FEB-1988; 88US-00160416.
XX PR 13-JUL-1989; 89US-00379076.
XX PR 13-JUL-1990; 90US-00553759.
XX PR 01-DEC-1992; 92US-00983647.
XX PA (GEO ) GEN HOSPITAL CORP.
XX PI Seed B, Aruffo A;
XX WPI; 1996-200279/20.
XX DR P-PSDB; AAR91433.
XX
XX Cloning of cDNA encoding cell surface antigen - useful for isolation of
XX diagnostic and therapeutic proteins.
XX Example 3; Fig 7A-7B; 79pp; English.
XX
XX A cDNA clone (AAT14707) codes for human CD28 antigen (AAR91433). It was
XX obtd. by constructing an expression library in COS cells using cDNA
XX derived from human T-cell tumour line HPB-ALL and vector pIH3M (see also
XX AAT14705), and panning of the library using antibody-coated plates. This
XX immunoselection cloning method, developed to clone genes for cell surface
XX antigens of human lymphocytes (see also AAT14703-04 and AAT14706-26), has
XX general appln. Cell surface antigens are obtd. for diagnostic and
XX therapeutic use. (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 1514 BP; 404 A; 359 C; 337 G; 414 T; 0 U; 0 Other;
XX
Query Match 68.0%; Score 20.4; DB 2; Length 1514;
Best Local Similarity 95.5%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 9 CTGTGTCAGGAGCGATAGGCTGC 30
XX 766 CGTGTCTCAGGAGCGATAGGCTGC 745
XX
RESULT 9
AAV63445/c
XX ID AAV63445 standard; cDNA; 1514 BP.
XX AC AAV63445;
XX DT 25-MAR-2003 (revised)
XX DT 07-JUN-1999 (first entry)
XX
DE Human CD28 cDNA.
DE KW CD28; cell surface antigen; human; lymphocyte; cloning; ss.
DE OS Homo sapiens.
DE FH Key Location/Qualifiers
DE FT CDS 100..762
DE FT sig_peptide 100..153
DE FT mat_peptide 154..759
DE FT /*tag= c
DE PN US5830731-A.
DE PD 03-NOV-1998.
DE PF 21-MAY-1997; 97US-00861205.
DE PR 25-FEB-1988; 88US-00160416.
DE PR 13-JUL-1989; 89US-00379076.
DE PR 23-MAR-1990; 90US-00498809.
DE PR 13-JUL-1990; 90US-00553759.
DE PR 01-DEC-1992; 92US-00983647.
DE PA (GEO ) GEN HOSPITAL CORP.
DE PI Seed B, Aruffo A;
DE WPI; 1998-609251/51.
DE DR P-PSDB; AAW80442.
DE
DE New cloning vector and poly:linker - based on existing sequences for
DE efficient cloning and expression of mammalian cDNA(s), especially human
DE lymphocyte antigenic sequences.
DE Example 3; Fig 7A-B; 75pp; English.
DE
DE This nucleotide sequence comprises human CD28 cDNA. The cDNA was isolated
DE from HPB-ALL human T-cell tumour cells using a novel method for cloning
DE cDNAs from mammalian expression libraries. The method is based on
DE transient expression of an antigen in eukaryotic cells and physical
DE selection of cells expressing the antigen by adhesion to an antibody-
DE coated substrate. The method is useful for the isolation and molecular
DE cloning of any protein which can be expressed and transported to the cell
DE surface membrane of a eukaryotic cell. It has been used to clone genes
DE (see AAV53442-63) encoding cell surface antigens associated with
DE mammalian lymphocytes such as CD1-53, ICAM, LFA-3, FCRIa, FCRIb, TL1a
DE and Leu8 (see AAW80440-55). CD28 polypeptide (see AAW80442) has been
DE expressed in transfected COS cells. The purified genes and proteins are
DE useful for immunodiagnostic and immunotherapeutic applications, including
DE the diagnosis and treatment of immune-mediated infections, diseases, and
DE disorders of animals, including humans. (Updated on 25-MAR-2003 to
DE correct PR field.)
DE
XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;
XX
Query Match 68.0%; Score 20.4; DB 2; Length 1514;
Best Local Similarity 95.5%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 9 CTGTGTCAGGAGCGATAGGCTGC 30
XX 766 CGTGTCTCAGGAGCGATAGGCTGC 745
XX
RESULT 10
AAV81203/c
XX ID AAV81203 standard; cDNA; 1514 BP.
XX AC AAV81203;
XX DT 25-MAR-2003 (revised)
XX DT 07-JUN-1999 (first entry)
XX
```

```

DT 10-MAY-1999 (first entry)
XX AC AAA50581;
DE DE Human CD28 cDNA.
XX DT
XX KW CD28; cell surface antigen; human; T cell antigen; T lymphocyte;
XX KW cDNA library; ss.
XX OS
XX FH Key Location/Qualifiers
XX FT CDS 100..762
XX FT /*tag= a
XX FT sig_peptide 100..153
XX FT mat_peptide 154..759
XX FT /*tag= c
XX PN US5849898-A.
XX PD 15-DEC-1998.
XX PF 07-JUN-1995; 95US-00485447.
XX PR 25-FEB-1988; 88US-00160416.
XX PR 13-JUL-1989; 89US-00379076.
XX PR 23-MAR-1990; 90US-00498809.
XX PR 13-JUL-1990; 90US-00553759.
XX PR 01-DEC-1992; 92US-00983647.
XX PA (GEO ) GEN HOSPITAL CORP.
XX PI Seed B, Oquendo C, Camerini D, Stamenkovic I, Stengelin S;
XX PI Aniot M, Lauffer L, Allen J, Simmons D, Aruffo A;
XX DR WPI; 1999-069813/06.
XX DR P-PSDB; AAW88451.
XX PT cDNA encoding human CD40 antigen - useful for cloning cDNA encoding cell
XX PT surface antigens, constructing cDNA libraries, expressing vectors for
XX PT expression in eukaryotic cells or their fragments.
XX PS Example 3; Fig 7A-B; 79pp; English.
XX CC This nucleotide sequence comprises human CD28 cDNA. The cDNA was isolated
XX CC from HPB-ALL human T-cell tumour cells using a novel method for cloning
XX CC cDNAs from mammalian expression libraries. The method is based on
XX CC transient expression of an antigen in eukaryotic cells and physical
XX CC selection of cells expressing the antigen by adhesion to an antibody-
XX CC coated substrate. The method is useful for the isolation and molecular
XX CC cloning of any protein which can be expressed and transported to the cell
XX CC surface membrane of a eukaryotic cell. It has been used to clone genes
XX CC (see AAV81198-220) encoding cell surface antigens such as CD1a, CD1b,
XX CC CD1c, CD2, CD6, CD7, CD13, CD14, CD16, CD19, CD20, CD22, CD26, CD27,
XX CC CD28, CD31, CD32a, CD32b, CD33, CD34, CD36, CD37, CD38, CD39, CD40,
XX CC CD43, CD44, CD53, ICAM, LFA-3, FCRIa, FCRIb, TLIa and Leu8 (see AAW86188
XX CC -62, AAW89151-52 and AAW88451). CD40 cDNA (see AAV81198) is specifically
XX CC claimed. CD28 polypeptide (see AAW88451) has been expressed in
XX CC transfected COS cells
XX SQ Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 68.0%; Score 20.4; DB 2; Length 1514;
Best Local Similarity 95.5%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTGAGGAGCGATAGGCTGC 30
DB 766 CGTGTGAGGAGCGATAGGCTGC 745

RESULT 11
AA50581/c
ID AAA50581 standard; cDNA; 1514 BP.

```

---

```

XX AC AAA50581;
XX DT 19-DEC-2000 (first entry)
XX DE Human cell surface antigen CD28 cDNA.
XX KW CD28; cell surface antigen; human; immunoselection; panning;
XX KW immunodiagnosis; diagnosis; immunotherapy; gene therapy; immune disorder;
XX KW infection; asthma; immune-complex disease; amyloidosis;
XX KW multiple sclerosis; parasitic disease; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 100..762
XX FT /*tag= a
XX FT sig_peptide 100..153
XX FT mat_peptide 154..759
XX FT /*tag= c
XX PN US6111093-A.
XX PD 29-AUG-2000.
XX PF 28-OCT-1998; 98US-00181612.
XX PR 25-FEB-1988; 88US-00160416.
XX PR 13-JUL-1989; 89US-00379076.
XX PR 23-MAR-1990; 90US-00498809.
XX PR 13-JUL-1990; 90US-00553759.
XX PR 01-DEC-1992; 92US-00983647.
XX PA (GEO ) GEN HOSPITAL CORP.
XX PI Stamenkovic I, Seed B;
XX DR WPI; 2000-586382/55.
XX DR P-PSDB; AAY96128.
XX PT Isolated nucleic acid molecule encoding the CD19 cell surface antigen,
XX PT useful for immunodiagnosis and immunotherapy of immune-mediated
XX PT infections or disorders, e.g. asthma, immune-complex disease, parasitic
XX PT diseases.
XX PS Example 3; Fig 7A-B; 75pp; English.
XX CC The present sequence is that of cDNA encoding human cell surface antigen
XX CC (CSA) CD28 (see AAY96128). The cDNA was derived from a human T-cell
XX CC tumour line HPS-ALL cDNA library using a new method for cloning CSA
XX CC cDNAs. The method is based upon transient expression of CSA in eukaryotic
XX CC cells and physical selection of cells expressing the antigen by adhesion
XX CC to (panning on) an antibody-coated substrate such as a culture dish. The
XX CC predicted amino acid sequence of CD28 suggests an integral membrane
XX CC protein with a single membrane-spanning hydrophobic domain terminating in
XX CC a 41-amino acid cytoplasmic domain. CSA nucleic acids isolated by the
XX CC method of the invention, and the proteins they encode, are useful for
XX CC immunodiagnostic and immunotherapeutic applications, including the
XX CC diagnosis and treatment of immune-mediated infections, diseases, and
XX CC disorders in animals, including humans. These disorders include asthma,
XX CC immune-complex disease, amyloidosis, parasitic diseases or multiple
XX CC sclerosis
XX SQ Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 68.0%; Score 20.4; DB 3; Length 1514;
Best Local Similarity 95.5%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTGAGGAGCGATAGGCTGC 30
DB 766 CGTGTGAGGAGCGATAGGCTGC 745

```

## RESULT 12

AAZ29325/c  
ID AAZ29325 standard; cDNA; 1514 BP.

XX AC AAZ29325;

XX AC AAZ29325;

XX 29-FEB-2000 (first entry)

XX Human CD28 receptor cDNA.

XX Human CD28; B7 molecule; immune response; cell surface receptor;  
KW Major histocompatibility complex; MHC classII; proton motor force;  
KW mitochondrial membrane potential; mitochondrial metabolism; cancer;  
KW autoimmune disease; neurodegenerative disorder; ds.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 100..762  
FT /\*tag= a  
FT /product= "Human CD28"

XX WO953953-A2.

XX 28-OCT-1999.

XX 30-MAR-1999; 99WO-US006874.

XX 17-APR-1998; 98US-0082250P.

XX 29-JUL-1998; 98US-0094519P.

XX 24-SEP-1998; 98US-0101580P.

XX (UYVE-) UNIV VERMONT.

XX Newell MK;

XX WPI; 2000-096773/08.

XX P-PSDB; AAY44294.

XX Use of cell surface and membrane characteristics for developing products  
PT for treating cancers, autoimmune diseases or neurodegenerative diseases.

XX Disclosure; Page 120-121; 123pp; English.

XX The present sequence encodes human CD28 receptor. CD28 is a homodimeric  
CC glycoprotein involved in a secondary signalling pathway in the activation  
CC of T-cell proliferation. B7 on nerve cells can interact with CD28 on the  
CC immune cell leading to immune cell activation. The regulation of cell  
CC surface expression of MHC classII and co-stimulatory molecule B7 can be  
CC manipulated by regulating the intracellular dissipation of proton motor  
CC force which can be assessed in terms of mitochondrial membrane potential.  
CC These methods can be used for regulating cell growth and division to  
CC control disease processes by manipulating mitochondrial metabolism and  
CC the expression of cell surface immune proteins. They can be used for  
CC treating diseases associated with excessive cellular division, aberrant  
CC differentiation, and premature cellular death, e.g. cancers, autoimmune  
CC diseases, neurodegenerative disorders etc

XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 68.0%; Score 20.4; DB 3; Length 1514;

Best Local Similarity 95.5%; Pred. No. 37;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30

Db 766 CGTGTACGAGCGATAGGCTGC 745

## RESULT 13

AAS03175/c

ID AAS03175 standard; cDNA; 1514 BP.

XX AAS03175;

XX 29-AUG-2001 (first entry)

XX Human lymphocyte cell surface antigen CD28 cDNA sequence.

XX Human; lymphocyte cell surface antigen; immune-mediated disease; CD28;  
KW infection; immune deficiency disorder; hypersensitivity; inflammation;  
KW systemic lupus erythematosus; platelet disorder; rheumatoid arthritis;  
KW transplant rejection; asthma; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 100..762  
FT /\*tag= a  
FT /product= "CD28 antigen"  
FT sig\_peptide 100..153  
FT /\*tag= b  
FT mat\_peptide 154..759  
FT /\*tag= c

XX US6218525-B1.

XX 17-APR-2001.

XX 01-DEC-1992; 92US-00983647.

XX 25-FEB-1988; 88US-00160416.

XX 13-JUL-1989; 89US-00379076.

XX 13-JUL-1990; 90US-00553759.

XX (GEO ) GEN HOSPITAL CORP.

XX Seed B, Aruffo A, Simmons D;

XX WPI; 2001-289848/30.

XX P-PSDB; AAU02437.

XX New recombinant DNA encoding CD28 useful for diagnosing and treating  
PT immune-mediated diseases, infections or disorders, e.g. systemic lupus  
PT erythematosus, asthma, transplant rejection, rheumatoid arthritis.

XX Example 3; Fig 7A-7B; 72pp; English.

XX The present sequence encoding for human lymphocyte cell surface antigen  
CC CD28 is used to obtain a new genetically engineered cDNA sequence  
CC encoding the CD28 amino acid extracellular domain sequence (amino acids 1  
CC -134 given in AAU02437) and/or comprising nucleotides 100-759, 154-555 or  
CC 154-759 of the CD28 cDNA sequence. Various human lymphocyte cell surface  
CC antigen cDNA sequences (AAS03172, AAS03173, AAS03175-AAS03195) are  
CC described in the present invention. The invention relates to a novel  
CC method of cloning cDNA encoding cell surface antigens and efficient  
CC construction of cDNA libraries. Also described are 2 expression vectors  
CC (AAS03171, AAS03174) which provide high level expression in eukaryotic  
CC host cells. The purified genes and proteins are useful for  
CC immunodiagnostic and immunotherapeutic applications, such as in the  
CC diagnosis and treatment of immune-mediated diseases, infections or  
CC disorders in animals and humans. Such diseases include immune deficiency  
CC diseases, diseases of immediate type of hypersensitivity, asthma,  
CC hypersensitivity pneumonitis, systemic lupus erythematosus, rheumatoid  
CC arthritis, acute and chronic inflammation, platelet disorders, plasma and  
CC other cell neoplasms, parasitic diseases, multiple sclerosis, Guillain-  
CC Barre syndrome and tissue and organ transplant rejection. The sequences  
CC can also be used to identify, isolate and purify other antibodies and  
CC antigens

XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;

Query Match 68.0%; Score 20.4; DB 4; Length 1514;

Best Local Similarity 95.5%; Pred. No. 37;

```
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30
Db 766 CGTGTACAGGAGCGATAGGCTGC 745

RESULT 14
ABA99035/c
ID ABA99035 standard; DNA; 1514 BP.
XX
AC ABA99035;
XX
XX 13-JUN-2002 (first entry)
XX
DE Human CD28 gene.
XX
KW Human; CD28; immune response; T cell; graft-versus-host disease; GVHD;
KW marrow transplant rejection; organ transplant rejection; psoriasis;
KW tissue transplant rejection; autoimmune disease; diabetes mellitus;
KW multiple sclerosis; rheumatoid arthritis; systemic lupus erythematosus;
KW systemic sclerosis; dermatomyositis; polymyositis; Sjogren's syndrome;
KW polyarteritis nodosa; vasculitis; sepsis; gene; ds.
XX
OS Homo sapiens.
XX
XX US2002006403-A1.
XX
XX 17-JAN-2002.
XX
XX 14-DEC-2000; 2000US-00738546.
XX
XX 14-DEC-1999; 99US-0170857P.
XX
XX (YUXX/) YU X.
XX (ANAS/) ANASETTI C.
XX
XX Yu X, Anasetti C;
XX
XX WPI; 2002-187719/24.
XX
XX Inhibiting an immune response in a subject for treating autoimmune
XX disease such as psoriasis, diabetes mellitus, multiple sclerosis,
XX rheumatoid arthritis, polymyositis, by administering anti-CD28 antibody.
XX
XX Disclosure; Page 28; 37pp; English.
XX
XX The sequence represents the human CD28 gene. The invention relates to a
XX novel method for inhibiting an immune response by administering an anti-
XX CD28 antibody. The method of the invention has immunosuppressive,
XX antipsoriatic, antidiabetic, antirheumatic, antineuropathic,
XX neuroprotective, dermatological, vasotropic, antiinflammatory,
XX antianaemic, hepatotropic, antiulcer, antibacterial and ophthalmological
XX activity. The method works to inhibit immune response by reversing or
XX blocking T cell activation. The method is useful for inhibiting an immune
XX response in a subject susceptible to graft-versus-host disease (GVHD),
XX marrow transplant rejection, organ transplant rejection or tissue
XX transplant rejection, or having autoimmune disease including psoriasis,
XX diabetes mellitus, multiple sclerosis, rheumatoid arthritis, systemic
XX lupus erythematosus, dermatomyositis, polymyositis,
XX Sjogren's syndrome, polyarteritis nodosa or vasculitis. The method is
XX also useful for treating sepsis, and other autoimmune diseases including
XX autoimmune hepatitis, autoimmune haemolytic anaemia, Behcet's disease,
XX myasthenia gravis, cirrhosis, uveitis, ulcerative colitis, and vitiligo
XX
XX Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;
XX
XX Query Match 68.0%; Score 20.4; DB 6; Length 1514;
XX Best Local Similarity 95.5%; Pred. No. 37;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30
Db 766 CGTGTACAGGAGCGATAGGCTGC 745

RESULT 16
ADI31990/c
ID ADI31990 standard; cDNA; 1514 BP.
XX
AC ADI31990;
XX
XX 17-JUN-2004 (first entry)
XX
DE Human cDNA #1316.
XX
XX Human; gene; ss; immunological response; immunopathological condition;
XX Crohn's disease; asthma; ulcerative colitis; hyper eosinophilia;
XX irritable bowel syndrome; osteoarthritis; rheumatoid arthritis;
XX acute monocytic leukaemia; antiinflammatory; antiasthmatic; antiulcer;
XX
```

KW osteopathic; antiarthritic; antirheumatic; cytostatic.  
XX Homo sapiens.  
XX US6607879-B1.  
XX 19-AUG-2003.  
XX 09-FEB-1998; 98US-00023655.  
XX 09-FEB-1998; 98US-00023655.  
XX (INCYTE) INCYTE CORP.  
XX Cocks BG, Stuart SG, Seilhamer JU;  
XX WPI; 2003-895307/82.  
XX A composition comprising a plurality of cDNAs, useful for detecting  
PT altered expression of genes in an immunological response or for  
PT diagnosing and treating an immunopathology, e.g. Crohn's disease, asthma  
PT or osteoarthritis.  
XX Claim 1; SEQ ID NO 1316; 50pp; English.  
XX The invention relates to a composition comprising a plurality of cDNAs  
CC for detecting the altered expression of genes in an immunological  
CC response. The invention also relates to a method of diagnosing or  
CC monitoring the treatment of an immunopathological condition in a sample,  
CC comprising obtaining nucleic acids from a sample, contacting the nucleic  
CC acids of the sample with an array comprising the plurality of cDNAs under  
CC conditions to form one or more hybridisation complexes, detecting the  
CC hybridisation complexes and comparing the levels of the detected  
CC hybridisation complexes with the level of hybridisation complexes  
CC detected in a non-diseased sample, where an altered level of the detected  
CC hybridisation complexes correlates with the presence of an  
CC immunopathological condition. Also disclosed are an expression profile  
CC comprising a microarray and a plurality of detectable complexes and a  
CC method for identifying a plurality of polynucleotide probes. The cDNAs  
CC are useful as hybridisable array elements in a microarray for monitoring  
CC the expression of target polynucleotides. The microarray can be used in  
CC the diagnosis of an immunopathology, such as Crohn's disease, asthma,  
CC ulcerative colitis, hyper eosinophilia, irritable bowel syndrome,  
CC osteoarthritis, rheumatoid arthritis or acute monocytic leukaemia, and in  
CC identifying agents for the treatment of the diseases. The microarray may  
CC also be used in drug discovery and development, toxicological and  
CC carcinogenicity studies, forensics or pharmacogenomics. The composition  
CC may also be used in purification of a subpopulation of mRNAs. cDNAs or  
CC genomic fragments. This sequence represents a human cDNA of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification but was obtained in electronic format directly  
CC from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
SQ Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;  
Query Match 68.0%; Score 20.4; DB 11; Length 1514;  
Best Local Similarity 95.5%; Pred. No. 37;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 9 CTTGTCCAGGAGCGATAGGCTGC 30  
Db 766 CGTGTCCAGGAGCGATAGGCTGC 745  
RESULT 17  
ADO49343/c  
ID ADO49343 standard; cDNA; 1514 BP.  
XX  
XX ADO49343;  
XX  
XX 15-JUL-2004 (first entry)  
DT Human CD28 antigen cDNA.  
DE

XX cell surface antigen; immune-mediated disorder; asthma;  
KW rheumatoid arthritis; multiple sclerosis; vasculitis; inflammation; ss;  
KW gene; human.  
XX Homo sapiens.  
XX US2004072283-A1.  
XX 15-APR-2004.  
XX 17-APR-2001; 2001US-00836544.  
XX 25-FEB-1988; 88US-00160416.  
PR 13-JUL-1989; 89US-00379076.  
PR 23-MAR-1990; 90US-00498809.  
PR 13-JUL-1990; 90US-00553759.  
PR 01-DEC-1992; 92US-00983647.  
XX (SEED/) SEED B.  
PA (ALLE/) ALLEN J.  
PA (ARUF/) ARUFFO A.  
PA (CAME/) CAMERINI D.  
PA (LAUF/) LAUFFER L.  
PA (QUE/) OQUENDO C.  
PA (SIMM/) SIMMONS D.  
PA (STAM/) STAMENKOVIC I.  
PA (STEN/) STENGELIN S.  
PA (AMIO/) AMIOT M.  
XX Seed B, Allen J, Aruffo A, Camerini D, Laufer L, Oquendo C;  
PI Simmons D, Stamenkovic I, Stengelin S, Amiot M;  
XX WPI; 2004-328571/30.  
DR P-PSDB; ADO49344.  
XX New cloning cDNA segments encoding cell surface antigens of human  
PT lymphocytes, useful in diagnosing and treating asthma, rheumatoid  
PT arthritis, multiple sclerosis, vasculitis and inflammation and  
PT infections.  
XX Example 3; Fig 7; 75pp; English.  
XX The invention relates to a cloned cDNA segment encoding a cell surface  
CC antigen selected from CD1a, CD1b, CD1c, CD2, CD6, CD7, CD13, CD14, CD16,  
CC CD19, CD20, CD22, CD26, CD27, CD31, CD32a, CD32b, CD33, CD34, CD36,  
CC CD37, CD38, CD39, CD40, CD43 and their functional derivatives. The cell  
CC surface antigens of human lymphocytes prepared from the cDNAs are useful  
CC in diagnostic and therapeutic utility in immune-mediated disorders  
CC (asthma, rheumatoid arthritis, multiple sclerosis, vasculitis and  
CC inflammation) and infections in mammals, including humans. The present  
CC sequence represents a human cell surface antigen cDNA.  
XX  
SQ Sequence 1514 BP; 404 A; 360 C; 337 G; 413 T; 0 U; 0 Other;  
Query Match 68.0%; Score 20.4; DB 12; Length 1514;  
Best Local Similarity 95.5%; Pred. No. 37;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 9 CTTGTCCAGGAGCGATAGGCTGC 30  
Db 766 CGTGTCCAGGAGCGATAGGCTGC 745  
RESULT 18  
AAN90607/c  
ID AAN90607 standard; cDNA; 1574 BP.  
XX  
XX AAN90607;  
XX  
XX 09-SEP-2004 (revised)  
DT 25-MAR-2003 (revised)  
DT 31-OCT-2002 (revised)

DT 20-DEC-1989 (first entry)  
 XX CD28 antigen cDNA.  
 DE  
 XX Tyrosine suppressor gene; ASV LTR; human cytomegalovirus AD169 enhancer;  
 KW HIV box; immunoselection; immune deficiency diseases; vasculitis;  
 KW systemic lupus erythematosus; rheumatoid arthritis; neoplasms; ss.  
 XX  
 OS Homo sapiens.  
 OS Unidentified.  
 XX  
 XX Key Location/Qualifiers  
 FH 100..819  
 FT /\*tag= a  
 FT 100..274  
 FT /\*tag= b  
 XX  
 XX EP330191-A.  
 PN  
 XX 30-AUG-1989.  
 PD  
 XX 23-FEB-1989; 89EP-00103127.  
 XX  
 XX 25-FEB-1988; 88US-00160416.  
 PR  
 XX (GEO) GEN HOSPITAL CORP.  
 XX  
 XX Seed B, Allen J, Aruffo A, Camerini D, Lauffer L, Oquendo CP;  
 PI Simmons D, Stamenkovi I;  
 XX WPI; 1989-250302/35.  
 DR P-PSDB; AAP91324.  
 DR  
 XX Rapid immuno:selection cloning - used to clone genes encoding cell  
 PT surface antigens associated with mammalian T lymphocytes.  
 PT  
 XX Disclosure; Fig 7; 69pp; English.  
 PS  
 XX This is used for cloning into a vector which transforms COS cells. The  
 CC vectors can be used to isolate any protein by immunoselection, and clones  
 CC are easy to manipulate. (Updated on 31-OCT-2002 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 CC Revised record issued on 09-SEP-2004 : Correction to keywords  
 CC  
 XX Sequence 1574 BP; 423 A; 375 C; 347 G; 429 T; 0 U; 0 Other;  
 SQ  
 Query Match 68.0%; Score 20.4; DB 1; Length 1574;  
 Best Local Similarity 95.5%; Pred. No. 38;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 9 CTTGTCTCAGGAGCGATAGGCTGC 30  
 Db 826 CGTGTCTCAGGAGCGATAGGCTGC 805  
 RESULT 19  
 AAT27652/c  
 ID AAT27652 standard; cDNA; 2405 BP.  
 XX  
 AC AAT27652;  
 XX  
 XX 24-AUG-1996 (first entry)  
 DT LFA-1-beta-CD28 chimeric receptor.  
 XX  
 DE LFA-1-beta-CD28 chimeric receptor.  
 XX  
 XX Human; chimeric receptor; LFA-1-beta; extracellular domain; CD28;  
 KW transmembrane region; intracellular domain; retrovirus; vector;  
 KW packaging cell culture; gene transfer; ICAM-1; selective activation;  
 KW cytotoxic T-lymphocyte; antitumour; immunotherapy; cancer; therapy;  
 XX virucide; ss.  
 XX  
 XX Homo sapiens.  
 OS

XX WO9613584-A1.  
 PN  
 XX 09-MAY-1996.  
 PD  
 XX 01-NOV-1995; 95WO-US014171.  
 PF  
 XX 01-NOV-1994; 94US-00332993.  
 PR  
 XX 07-JUN-1995; 95US-00480577.  
 PR  
 XX (TARG-) TARGETED GENETICS CORP.  
 PA  
 XX Feldhaus AL, Jones LA;  
 PI WPI; 1996-239497/24.  
 DR  
 XX TH independent cytotoxic T lymphocytes useful in tumour or viral therapy  
 PT - are activated by adhesion to antigen displaying cells.  
 PT  
 XX Example 1; Page 57-58; 82pp; English.  
 PS  
 XX The sequence is an XbaI-Sall fragment encoding a chimeric receptor,  
 CC composed of the extracellular region (XR) of leukocyte function-  
 CC associated antigen-1 beta-chain (LFA-1-beta, of the integrin family),  
 CC fused via the transmembrane region (TM) of CD28 to the intracellular  
 CC region (IC28) of CD28. TM and IC28 coding sequences (bp 556-767) are  
 CC amplified by PCR from human CD28 cDNA, and the XR coding sequence (bp 1-  
 CC 2072) is amplified by PCR from human LFA-1-beta cDNA. The sequences are  
 CC cloned to form plasmid pSKLFA-1-beta/CD28, and subcloned in HyTK-CMV-Cat  
 CC to form HyTK-CMV-LFA-1-beta/CD28. This retrovirus vector is introduced  
 CC into a psi-2 ecotropic packaging cell culture to give a stable PA317 cell  
 CC line producing high titers of infectious particles, which are then used  
 CC to infect T-lymphocytes. The resulting cytotoxic T-lymphocytes (CTLs) are  
 CC selectively activatable, and have a lessened dependence on helper T-  
 CC lymphocytes (TH cells) and/or growth factors. The chimeric receptor binds  
 CC to the LFA-1 ligand (intercellular adhesion molecule-1, ICAM-1 or CD54)  
 CC but IC28 transmits the proliferative signal of CD28. The CTLs mediate  
 CC lysis of a target cell bearing a cognate antigen, and are useful in  
 CC directed cancer immunotherapy, and in treatment of viral infections  
 XX  
 XX Sequence 2405 BP; 503 A; 736 C; 739 G; 427 T; 0 U; 0 Other;  
 SQ  
 Query Match 68.0%; Score 20.4; DB 2; Length 2405;  
 Best Local Similarity 95.5%; Pred. No. 40;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 9 CTTGTCTCAGGAGCGATAGGCTGC 30  
 Db 2398 CGTGTCTCAGGAGCGATAGGCTGC 2377  
 RESULT 20  
 AAT27651/c  
 ID AAT27651 standard; cDNA; 3585 BP.  
 XX  
 AC AAT27651;  
 XX  
 XX 24-AUG-1996 (first entry)  
 DT LFA-1-alpha-CD28 chimeric receptor.  
 DE  
 XX Human; chimeric receptor; LFA-1-alpha; extracellular domain; CD28;  
 KW transmembrane region; intracellular domain; retrovirus; vector;  
 KW packaging cell culture; gene transfer; ICAM-1; selective activation;  
 KW cytotoxic T-lymphocyte; antitumour; immunotherapy; cancer; therapy;  
 XX virucide; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO9613584-A1.  
 PN  
 XX 09-MAY-1996.  
 PD



PF 01-NOV-1995; 95WO-US014171.  
XX  
PR 01-NOV-1994; 94US-00332993.  
PR 07-JUN-1995; 95US-00480577.  
XX  
XX (TARG-) TARGETED GENETICS CORP.  
XX  
XX Feldhaus AL; Jones LA;  
PI  
XX WPI; 1996-239497/24.  
DR  
XX  
XX TH independent cytotoxic T lymphocytes useful in tumour or viral therapy  
PT - are activated by adhesion to antigen displaying cells.  
PT  
XX  
PS Example 1; Page 55-57; 82pp; English.  
XX  
CC The sequence is an XbaI-KpnI fragment encoding a chimeric receptor,  
CC composed of the extracellular region (XR) of leukocyte function-  
CC associated antigen-1 alpha-chain (LFA-1-alpha, of the integrin family),  
CC fused via the transmembrane region (TM) of CD28 to the intracellular  
CC region (IC28) of CD28. TM and IC28 coding sequences (bp 556-767) are  
CC amplified by PCR from human CD28 cDNA, and the XR coding sequence (bp 1-  
CC 3358) is amplified by PCR from human LFA-1- alpha cDNA. The sequences are  
CC cloned to form plasmid pSKLFA-1- alpha/CD28, and subcloned in tgSpIusNeo  
CC -CMV to form tgSpIusNeo- CMV-LFA-1-alpha/CD28. This retrovirus vector is  
CC introduced into a psi-2 ecotropic packaging cell culture to give a stable  
CC PA317 cell line producing high titers of infectious particles, which are  
CC then used to infect T-lymphocytes. The resulting cytotoxic T-lymphocytes  
CC (CTLs) are selectively activatable, and have a lessened dependence on  
CC helper T-lymphocytes (TH cells) and/or growth factors. The chimeric  
CC receptor binds to the LFA-1 ligand (intercellular adhesion molecule-1,  
CC ICAM-1 or CD54) but IC28 transmits the proliferative signal of CD28. The  
CC CTLs mediate lysis of a target cell bearing a cognate antigen, and are  
CC useful in directed cancer immunotherapy, and in treatment of viral  
CC infections  
XX  
SQ Sequence 3585 BP; 808 A; 981 C; 1011 G; 785 T; 0 U; 0 Other;  
Query Match 68.0%; Score 20.4; DB 2; Length 3585;  
Best Local Similarity 95.5%; Pred. No. 42;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 9 CTGTGTCAGGAGCGATAGGCTGC 30  
DB 3578 CGTGTGTCAGGAGCGATAGGCTGC 3557  
RESULT 21  
AAL49546/c  
ID AAL49546 standard; DNA; 3803 BP.  
XX  
XX AAL49546;  
AC  
XX  
XX 27-NOV-2002 (first entry)  
DT  
XX  
DE Human CD28 coding sequence.  
XX  
XX Human; CD28; immune response; signaling; diabetes; autoimmune disease;  
KW immunosuppressive; dermatological; anti-inflammatory; antithyroid;  
KW antirheumatic; antiarthritic; neutropic; allergy; rheumatoid arthritis;  
KW systemic lupus erythematosus; myasthenia gravis; autoimmune thyroiditis;  
KW vitiligo; alopecia; inflammatory bowel disease; Addison's disease;  
KW Graves disease; haemolytic anaemia; Sjogren's syndrome; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 222. .884  
CDS /\*tag= a  
FT /product= "CD28"  
FT /transl\_except= (pos:867. .872,aa:Ala)  
XX  
XX WO200266059-A2.

XX 29-AUG-2002.  
PD  
XX 15-FEB-2002; 2002WO-US004772.  
PF  
XX 16-FEB-2001; 2001US-0269756P.  
PR  
XX (GENY ) GENETICS INST INC.  
PA  
XX O'hara RM, Nagelin AM;  
PI  
XX WPI; 2002-674899/72.  
DR P-PSDB; AAC19101.  
XX  
PT Downmodulating immune responses by blocking CD28-mediated signaling,  
PT useful for preventing and/or treating autoimmune disorders such as  
PT diabetes, allergic reactions, graft versus host disease, systemic lupus  
PT erythematosus.  
XX  
PS Disclosure; Page 60-61; 61pp; English.  
XX  
CC The present invention relates to a method of therapeutically  
CC downmodulating an autoimmune response or an ongoing autoimmune response,  
CC comprising administering an antigen binding portion of an anti-CD28  
CC antibody that blocks signaling via CD28 to the subject so that an  
CC autoimmune response or an ongoing autoimmune response in the subject is  
CC downmodulated. The methods are useful in therapeutically and  
CC prophylactically downmodulating the immune response in subjects having  
CC autoimmune disorders such as diabetes, allergy and allergic reactions,  
CC transplantation rejection, graft versus host disease, systemic lupus  
CC erythematosus, rheumatoid arthritis, myasthenia gravis, autoimmune  
CC thyroiditis, vitiligo, alopecia, inflammatory bowel disease, Addison's  
CC disease, Graves disease, haemolytic anaemia and Sjogren's syndrome. The  
CC present sequence is the human CD28 coding sequence  
XX  
SQ Sequence 3803 BP; 1076 A; 766 C; 880 G; 1081 T; 0 U; 0 Other;  
Query Match 68.0%; Score 20.4; DB 6; Length 3803;  
Best Local Similarity 95.5%; Pred. No. 43;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 9 CTGTGTCAGGAGCGATAGGCTGC 30  
DB 888 CGTGTGTCAGGAGCGATAGGCTGC 867  
RESULT 22  
ABV75351/c  
ID ABV75351 standard; DNA; 3804 BP.  
XX  
XX ABV75351;  
AC  
XX  
XX 07-MAR-2003 (first entry)  
DT  
XX  
DE Human CD28 protein encoding DNA.  
XX  
XX T-cell; cell proliferation; cell survival; cytostatic; antiallergic;  
KW immunostimulant; immunosuppressive; gene therapy; CD28; human; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 223. .885  
CDS /\*tag= a  
FT /product= "CD28"  
FT  
XX  
XX WO200290518-A2.  
PN  
XX  
XX 14-NOV-2002.  
PD  
XX  
XX 10-MAY-2002; 2002WO-US014843.  
PF  
XX  
XX 10-MAY-2001; 2001US-0290097P.  
PR

XX PA (UNIW ) UNIV WASHINGTON.  
XX PI Green JM, Shaw AS;  
XX DR WPI; 2003-111968/10.  
XX DR P-PSDB; ABB82721.  
XX PT Selectively modulating T-cell survival or proliferation, useful for  
PT treating autoimmune diseases, comprises contacting T-cells expressing  
PT CD28 protein with agents that modulate activity of CD28 survival or  
PT proliferation motif.  
XX PS Disclosure; Page 115-117; 119pp; English.  
XX CC The invention relates to selectively modulating T-cell survival relative  
CC to T-cell proliferation, or T-cell proliferation relative to T-cell  
CC survival. The method involves contacting a T-cell expressing a CD28  
CC protein with an agent that selectively modulates the activity of a CD28  
CC survival or proliferation motif relative to the activity of a CD28  
CC proliferation or survival motif, respectively, to selectively modulate  
CC survival or proliferation of the T-cell. The method is useful in  
CC selectively modulating survival or proliferation of T-cells in a subject.  
CC The method may also be used to prevent or treat disorders associated with  
CC aberrant cell proliferation or survival, e.g. cancer, transplant-  
CC associated disorders, allergic diseases, graft-versus-host disease,  
CC lymphoproliferative disorders or autoimmune diseases. The present  
CC sequence represents a DNA encoding the human CD28 protein  
XX SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;  
Query Match 68.0%; Score 20.4; DB 8; Length 3804;  
Best Local Similarity 95.5%; Pred. No. 43;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 9 CTTGTCAGGAGCGATAGGCTGC 30  
Db 889 CGTGTCTCAGGAGCGATAGGCTGC 868  
RESULT 23  
ADAO2991/c  
ID ADAO2991 standard; cDNA; 3804 BP.  
XX AC ADAO2991;  
XX DT 06-NOV-2003 (first entry)  
XX DE Human CD28 carcinoma associated cDNA, SEQ ID NO:1509.  
XX KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ss.  
XX OS Homo sapiens.  
XX PN WO2003057146-A2.  
XX PD 17-JUL-2003.  
XX PF 26-DEC-2002; 2002WO-US041414.  
XX PR 26-DEC-2001; 2001US-00035832.  
XX PA (SAGR-) SAGRES DISCOVERY.  
XX PI Morris DW;  
XX FI WPI; 2003-587068/55.  
XX DR New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX PT

PS CC The invention relates to recombinant carcinoma associated (CA) nucleic  
XX acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or  
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed human CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;  
Query Match 68.0%; Score 20.4; DB 9; Length 3804;  
Best Local Similarity 95.5%; Pred. No. 43;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 9 CTTGTCAGGAGCGATAGGCTGC 30  
Db 889 CGTGTCTCAGGAGCGATAGGCTGC 868  
RESULT 24  
ADB72729/c  
ID ADB72729 standard; mRNA; 3804 BP.  
XX AC ADB72729;  
XX DT 04-DEC-2003 (first entry)  
XX DE Human CD28 mRNA.  
XX KW human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
KW cancer; neoplasm; adenocarcinoma; sarcoma.  
XX OS Homo sapiens.  
XX PN WO2003008583-A2.  
XX PD 30-JAN-2003.  
XX PF 26-DEC-2001; 2001WO-US051291.  
XX PR 02-MAR-2001; 2001US-00798586.  
XX PR 23-OCT-2001; 2001US-00004113.  
XX PR 08-NOV-2001; 2001US-00052482.  
XX PR 30-NOV-2001; 2001US-00997722.  
XX PR 20-DEC-2001; 2001US-00034650.  
XX PA (SAGR-) SAGRES DISCOVERY.  
XX PI Morris DW, Engelhard EK;  
XX FI WPI; 2003-239337/23.  
XX DR New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
XX PT  
PS Claim 1; SEQ ID NO 557; 2304pp; English.

xx The invention relates to a novel recombinant nucleic acid comprising a CC nucleotide sequence selected from any of the 660 sequences fully defined CC in the specification. A polynucleotide of the invention has cytostatic CC activity, and may have a use in gene therapy, or in a vaccine. The CC recombinant nucleic acids and polypeptides are useful for treating CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and CC sarcomas. The present sequence represents a human mRNA of the invention. CC

SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;  
 Query Match 68.0%; Score 20.4; DB 10; Length 3804;  
 Best Local Similarity 95.5%; Pred. No. 43;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**Qy** 9 CTTGT CAGGAGCGATAGGCTGC 30  
| | | | | | | | | |  
**Db** 889 CGTGT CAGGAGCGATAGGCTGC 868

RESULT 25  
ADC85471/C  
ID ADC85471 standard: DNA: 3804 BP.

01-JAN-2004 (first entry)

Human Cd28 mRNA sequence.

Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA; secreted; transmembrane; intracellular; ds.

OS Homo sapiens.

PN WO2003045230-A2.

05-JUN-2003.

02-DEC-2002; 2002WO-US038582.

AA  
PR 30-NOV-2001; 2001US-00997722.

PA (SAGR-) SAGRES DISCOVERY.

PI Morris DW. Engelhard EK:

WPI: 2003-513603/48.

New recombinant nucleic acid comprising a nucleotide sequence of any of the carcinoma-associated (CA) genes, useful for screening for drug candidates for diagnosing or treating carcinomas.

PS claim 1: SEO ID NO 257: 983pp: English.

The invention relates to a recombinant nucleic acid comprising a nucleotide sequence selected from any of the fully defined carcinoma-associated (CA) genes from the 50 tables given in the specification. The CA proteins are secreted, transmembrane or intracellular proteins. The recombinant nucleic acids are useful for screening for drug candidates for diagnosing or treating carcinomas. Sequences given in APC85215-APC85514 represent CA genes of the invention.

[illegible]

RESULT 26  
ADM74586/0

XX  
AC ADM74586:

01-JUL-2004 (first entry)

DE Human carcinoma associated (CA) nucleic acid #128.

Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ss;  
KW  
carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;  
KW  
cytostatic.  
KW

Homo sapiens.

PN US2004072154-A1.

15-APR-2004.

PF 30-NOV-2001; 2001US-00997722.

PR 22-DEC-2000; 2000US-00747377.

PK 02-MAR-2001; 2001007; 38588;  
XX

PA (MORR/) MORRIS D W.  
PA (ENGE/) ENGELHARD E K.

PI Morris DW, Engelhard EK;

DR WPI; 2004-328562/30.

PT New carcinoma associated gene or protein, useful for preparing a  
PT composition for diagnosing or treating carcinoma e.g., leukemia or  
PT lymphoma.

PS Claim 1: SEO ID NO 257; 29pp: English.

The invention relates to new recombinant nucleic acids. The invention also relates to a host cell comprising a recombinant nucleic acid or expression vector, an expression vector comprising a recombinant nucleic acid, a recombinant protein, a method of screening for drug candidates, a method of screening for a bioactive agent capable of binding to a carcinoma associated protein (CAP) encoded by a nucleotide sequence, a method of screening for a bioactive agent capable of modulating the activity of a CAP, a method of evaluating the effect of a candidate carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting the activity of a CAP, a method of treating carcinomas, a method of neutralising the effect of a CAP and a method of diagnosing carcinoma or propensity to carcinoma. A method of evaluating the effect of a candidate carcinoma drug comprises administering the drug to a patient, removing a cell sample from the patient and determining alterations in the expression or activation of a gene comprising the nucleotide sequence. A method of diagnosing carcinoma comprises determining the expression of one or more genes comprising the nucleic acid sequence in a first tissue type of a first individual and comparing the expression of the gene from a second normal tissue type from the first individual or a second unaffected individual, where a difference in the expression indicates that the first individual has carcinoma. A method of inhibiting the activity of a CAP comprises binding an inhibitor to the CAP. Treating carcinomas comprises administering to a patient an inhibitor of CAP. Neutralising the effect of a CAP comprises contacting an agent specific for the CAP. The polypeptide specifically binds to the protein encoded by the nucleic acid. It comprises an antibody that specifically binds to the protein encoded by the nucleic acid. The nucleic acids are useful for preparing a composition for diagnosing or treating carcinoma e.g., leukaemia or lymphoma. This sequence represents a human carcinoma associated (CA) nucleic acid of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at [scddata.uspto.gov/sequence.html](http://scddata.uspto.gov/sequence.html).

```

SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;
Query Match 68.0%; Score 20.4; DB 12; Length 3804;
Best Local Similarity 95.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30
Db 889 CGTGTACGAGCGATAGGCTGC 868

RESULT 27
ADN05855/c
ID ADN05855 standard; cDNA; 3804 BP.
XX AC
XX ADN05855;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic cDNA sequence #1159.
XX KW ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX KW P-PSDB; ADN05856.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT;
PI Wu TD;
XX WPI; 2004-305105/28.
XX DR P-PSDB; ADN05856.
XX DR
XX PT New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX PS Claim 1; SEQ ID NO 2250; 3069pp; English.
XX CC The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polynucleotides of the invention.
XX SQ Sequence 3804 BP; 1076 A; 766 C; 880 G; 1082 T; 0 U; 0 Other;
Query Match 68.0%; Score 20.4; DB 12; Length 3804;
Best Local Similarity 95.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30
Db 889 CGTGTACGAGCGATAGGCTGC 868

RESULT 28
AAD21976/c
ID AAD21976 standard; DNA; 3806 BP.
XX AC
XX AAD21976;
XX AC
XX DT 12-FEB-2002 (first entry)
XX DE Human CD28 DNA.
XX
```

```

XX KW Human; cytotoxic T lymphocyte antigen-4; CTLA-4; CD28; AIDS; therapy;
KW Acquired Immune Deficiency Syndrome; neuroprotective; dermatological;
KW immune response; organ transplantation; autoimmune disease; allergy; SLE;
KW systemic lupus erythematosus; multiple sclerosis; tumour vaccination;
KW immunodeficiency disease; DiGeorge Syndrome; cancer; ds.
XX OS Homo sapiens.
XX FH Key
XX CDS 223..885
XX FT /tag= a
XX FT /product= "Human CD28 protein"
XX PN WO200179300-A1.
XX PD 25-OCT-2001.
XX PF 12-APR-2001; 2001WO-US012275.
XX PR 12-APR-2000; 2000US-0196851P.
XX PA (GEMY ) GENETICS INST INC.
XX PI Bluestone JA, Collins M, Whitters M, Griffin M, Kranz D;
XX DR WPI; 2002-017603/02.
XX DR P-PSDB; AAB13721.
XX PT New construct for downmodulating immune response in a subject, has
PT exposed surface attached with antigen-binding portion of antibody that
PT binds to cytotoxic T lymphocyte antigen-4 and major histocompatibility
PT molecule.
XX PS Disclosure; Page 93-94; 98pp; English.
XX CC The invention relates to a construct for downmodulating immune response
CC in a subject. The construct comprises an exposed surface attached with an
CC antigen-binding portion of an antibody that binds to a cytotoxic T
CC lymphocytic antigen (CTLA)-4 or CD28 expressed on T-cell of the subject,
CC and a major histocompatibility complex (MHC) molecule. The construct is
CC useful for treating an individual afflicted with a disease or disorder
CC that would benefit from downregulation of immune response. Downmodulation
CC of the immune response is useful to downmodulate the immune response in
CC situations of tissue, skin and organ transplantation, graft-versus-host
CC disease or in autoimmune diseases such as systemic lupus erythematosus
CC and multiple sclerosis. The construct is useful for inhibiting immune
CC cell activation and prevents production of autoantibodies or cytokines
CC which may be involved in disease process. Inhibition of immune cell
CC activation is useful in the treatment of allergy and allergic reactions
CC e.g., by inhibiting IgE production. The construct is also useful for
CC enhancing desirable immune response in a subject e.g., in situations of
CC tumour vaccination, in viral immunity or in immunodeficiency diseases
CC such as AIDS and DiGeorge Syndrome. Induction/enhancement of immune cell
CC function results in increased tumour destruction in cancer patients. The
CC construct is also useful in treating infectious diseases. The present
CC sequence is human CD28 DNA
XX SQ Sequence 3806 BP; 1079 A; 765 C; 881 G; 1081 T; 0 U; 0 Other;
Query Match 68.0%; Score 20.4; DB 6; Length 3806;
Best Local Similarity 95.5%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTCAGGAGCGATAGGCTGC 30
Db 889 CGTGTACGAGCGATAGGCTGC 868

RESULT 29
ADA02990/c
ID ADA02990 standard; DNA; 51365 BP.
XX
```

AC ADA02990;  
 XX  
 DT 06-NOV-2003 (first entry)  
 XX  
 DE Human CD28 carcinoma associated gene, SEQ ID NO:1508.  
 XX  
 KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003057146-A2.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 26-DEC-2002; 2002WO-US041414.  
 XX  
 PR 26-DEC-2001; 2001US-00035832.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW;  
 XX  
 XX WPI; 2003-587068/55.  
 DR  
 XX New recombinant nucleic acid encoding carcinoma associated protein,  
 XX useful for preparing compositions for treating carcinomas.  
 PT  
 PT Claim 1; SEQ ID NO 1508; 245pp; English.  
 XX  
 PS The invention relates to recombinant carcinoma associated (CA) nucleic  
 XX acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 51365 BP; 15176 A; 10059 C; 10474 G; 15656 T; 0 U; 0 Other;  
 SQ Query Match 68.0%; Score 20.4; DB 9; Length 51365;  
 Best Local Similarity 95.5%; Pred. No. 62;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 9 CTTGTCAGGAGCGATAGGCTGC 30  
 DB 38448 CGTGTGAGGAGCGATAGGCTGC 38427  
 RESULT 30  
 ID ADB72728/c  
 AD B72728 standard; DNA; 51365 BP.  
 XX  
 AC ADB72728;  
 XX  
 XX 04-DEC-2003 (first entry)  
 DT  
 XX

DE Human CD28 gene.  
 XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003008583-A2.  
 XX  
 PD 30-JAN-2003.  
 XX  
 PF 26-DEC-2001; 2001WO-US051291.  
 XX  
 PR 02-MAR-2001; 2001US-00798586.  
 PR 23-OCT-2001; 2001US-00004113.  
 PR 08-NOV-2001; 2001US-00052482.  
 PR 30-NOV-2001; 2001US-00997722.  
 PR 20-DEC-2001; 2001US-00034650.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.  
 XX  
 PI Morris DW, Engelhard EK;  
 XX  
 XX WPI; 2003-239337/23.  
 DR  
 XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
 PT  
 PS Claim 1; SEQ ID NO 556; 2304pp; English.  
 XX  
 CC The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a human gene of the invention.  
 XX  
 SQ Sequence 51365 BP; 15176 A; 10059 C; 10474 G; 15656 T; 0 U; 0 Other;  
 Query Match 68.0%; Score 20.4; DB 10; Length 51365;  
 Best Local Similarity 95.5%; Pred. No. 62;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 9 CTTGTCAGGAGCGATAGGCTGC 30  
 DB 38448 CGTGTGAGGAGCGATAGGCTGC 38427  
 RESULT 31  
 ADC85470/c  
 ID ADC85470 standard; DNA; 51365 BP.  
 XX  
 AC ADC85470;  
 XX  
 XX 01-JAN-2004 (first entry)  
 DT  
 XX Human Cd28 genomic sequence.  
 DE  
 XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;  
 KW secreted; transmembrane; intracellular; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003045230-A2.  
 XX  
 PD 05-JUN-2003.  
 XX  
 PF 02-DEC-2002; 2002WO-US038582.  
 XX  
 PR 30-NOV-2001; 2001US-00997722.  
 XX  
 PA (SAGR-) SAGRES DISCOVERY.









XX BBebington CR, Lawson ADG, Weir ANC, Finney HM;  
 XX WPI; 1997-351052/32.  
 XX New DNA systems for activating cells - comprising DNA coding for a  
 PT chimeric receptor comprising 2 or more different cytoplasmic signalling  
 PT components.  
 PS Example 1; Fig 3; 90pp; English.  
 XX Primers P3301, P3302, P3303, P3304, P3305 and P3306 (see AAT90527-32)  
 CC were used for the PCR assembly of a human T cell receptor zeta-CD28  
 CC fusion cassette. The 3' end of zeta, starting at the native StyI site,  
 CC and the intracellular component of human CD28 were PCR assembled such  
 CC that the zeta stop codon was removed and an in-frame fusion protein would  
 CC be translated. The PCR product was subcloned into pluscript SK+  
 CC containing a CMV01 TCR zeta chimeric receptor construct, replacing the 3'  
 CC end of zeta. The zeta-CD28 fusion cassette can be utilised in novel  
 CC chimeric genes (see AAT90510 and AAT90513) encoding recombinant chimeric  
 CC receptors (see AAW26647 and AAW26650) useful in cell activation  
 CC processes, e.g. for the treatment of cancer  
 XX Sequence 64 BP; 16 A; 25 C; 10 G; 13 T; 0 U; 0 Other;  
 SQ

Query Match 64.7%; Score 19.4; DB 2; Length 64;  
 Best Local Similarity 95.2%; Pred. No. 67;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 10 TTCTCAGGAGCGATAGGCTGC 30  
 DB 58 TTCTCAGGAGCGATAGGCTGC 38

RESULT 39  
 ADA69416  
 ID ADA69416 standard; DNA; 852 BP.  
 XX AC ADA69416;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Rice gene, SEQ ID 2739.  
 XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
 XX gene; ds.  
 XX OS Oryza sativa.  
 XX PN WO2003000898-A1.  
 XX PD 03-JAN-2003.  
 XX PF 22-JUN-2001; 2001WO-IB001105.  
 XX PR 22-JUN-2001; 2001WO-IB001105.  
 XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 XX WPI; 2003-175290/17.  
 XX Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX Claim 6; SEQ ID NO 2739; 899pp; English.  
 XX The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1

CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX Sequence 852 BP; 219 A; 230 C; 267 G; 136 T; 0 U; 0 Other;  
 SQ

Query Match 64.7%; Score 19.4; DB 8; Length 852;  
 Best Local Similarity 79.3%; Pred. No. 98;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 2 CGAGGATCTTGTCCAGGAGCGATAGGCTGC 30  
 DB 417 CCAGGATCTCTCCAGGAGCGAGAGCTGC 445

RESULT 40  
 AAS81566  
 ID AAS81566 standard; CDNA; 1156 BP.  
 XX AC AAS81566;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE DNA encoding novel human diagnostic protein #17370.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX P-PSDB; ABG17379.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 1; SEQ ID NO 17370; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

```
Query Match      64.7%; Score 19.4; DB 5; Length 1156;
Best Local Similarity 79.3%; Pred. No. 1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

**Qy**            1 TCGAGGATCTTGT CAGGAGCGATAGGCTG 29  
               |||||  
**Db**            824 TCGAGGTTCTTCTGAGGAACGATGCGCTG 852

Search completed: January 7, 2005, 11:40:49  
Job time : 230.421 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 54.2105 Seconds  
(without alignments)  
393.349 Million cell updates/sec

Title: US-09-786-502A-8  
Perfect score: 30  
Sequence: 1 tcgagagcttgcaggagcgataggctgc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCUTUS\_COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.4	68.0	875	3	US-08-549-846-1
C 2	20.4	68.0	1064	2	US-08-529-878B-40
C 3	20.4	68.0	1514	4	US-09-023-655-1316
C 4	20.4	68.0	2405	3	US-08-549-846-3
C 5	20.4	68.0	3585	3	US-08-549-846-2
C 6	19.3	63.3	20	2	US-08-529-878B-18
C 7	19.3	63.3	1189	4	US-09-716-129-46
C 8	19.3	63.3	4758	4	US-09-774-528-129
C 9	18.0	60.0	29	3	US-08-812-946A-7
C 10	18.0	60.0	268	4	US-09-270-767-1027
C 11	18.0	60.0	768	4	US-09-270-767-16309
C 12	18.0	60.0	246	4	US-09-270-767-2215
C 13	18.0	60.0	746	4	US-09-270-767-17497
C 14	18.0	60.0	1498	4	US-09-270-767-15260
C 15	17.8	59.3	774	4	US-09-181-339-4
C 16	17.8	59.3	1974	4	US-09-221-017B-726
C 17	17.4	58.0	1758	4	US-09-252-991A-7065
C 18	17.4	58.0	2910	4	US-09-252-991A-7009
C 19	17.4	58.0	3186	4	US-09-252-991A-7120
C 20	17.2	57.3	351	4	US-09-513-999C-25301
C 21	17.2	57.3	534	4	US-09-270-767-4139
C 22	17.2	57.3	534	4	US-09-270-767-19421
C 23	17.2	57.3	939	4	US-09-489-039A-3420
C 24	17.2	57.3	1492	4	US-09-741-238-11
C 25	17.2	57.3	1692	4	US-09-600-991-3
C 26	17.2	57.3	1709	4	US-09-600-991-21
C 27	17.2	57.3	2021	1	US-07-882-925A-3

C 28	17.2	57.3	2021	1	US-08-184-012C-3	Sequence 3, Appli
C 29	17.2	57.3	2136	4	US-09-600-991-19	Sequence 19, Appl
C 30	17.2	57.3	2136	4	US-09-601-040A-11	Sequence 11, Appl
C 31	17.2	57.3	2188	1	US-07-882-925A-4	Sequence 4, Appli
C 32	17.2	57.3	2188	1	US-08-184-012C-4	Sequence 4, Appli
C 33	17.2	57.3	2216	2	US-08-666-082B-2	Sequence 2, Appli
C 34	17.2	57.3	2219	1	US-07-882-925A-1	Sequence 1, Appli
C 35	17.2	57.3	2219	1	US-07-882-925A-2	Sequence 2, Appli
C 36	17.2	57.3	2219	1	US-08-184-012C-1	Sequence 1, Appli
C 37	17.2	57.3	2219	1	US-08-184-012C-2	Sequence 2, Appli
C 38	17.2	57.3	2232	1	US-08-334-177-1	Sequence 1, Appli
C 39	17.2	57.3	2232	5	PCT-US95-13830-1	Sequence 1, Appli
C 40	17.2	57.3	2262	1	US-07-882-925A-7	Sequence 7, Appli
C 41	17.2	57.3	2262	1	US-08-184-012C-7	Sequence 7, Appli
C 42	17.2	57.3	2450	4	US-09-620-312D-336	Sequence 336, App
C 43	17.2	57.3	2513	4	US-09-620-312D-337	Sequence 337, App
C 44	17.2	57.3	7011	1	US-08-306-691B-42	Sequence 42, Appl
C 45	17	56.7	1013	4	US-08-930-830B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-549-846-1/c  
; Sequence 1, Application US/08549846  
; Patent No. 6083751  
; GENERAL INFORMATION:  
; APPLICANT: FELDHAUS, ANDREW L.  
; TITLE OF INVENTION: CHIMERIC RECEPTORS FOR THE GENERATION OF  
; TITLE OF INVENTION: SELECTIVITY-ACTIVATABLE T(h)-INDEPENDENT CYTOTOXIC T CELLS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/549,846  
; FILING DATE: 01-NOV-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DYLAN, TYLER  
; REGISTRATION NUMBER: 37,612  
; REFERENCE/DOCKET NUMBER: 22627-20013.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141 MRSNFOERS SFO  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 875 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-549-846-1

Query Match 68.0%; Score 20.4; DB 3; Length 875;  
Best Local Similarity 95.5%; Pred. No. 2.5;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTGAGGAGCGATAGGCTGC 30  
DB 868 CGTGTGAGGAGCGATAGGCTGC 847

## RESULT 2

```

US-08-529-878B-40/c
; Sequence 40, Application US/08529878B
; Patent No. 5932556
; GENERAL INFORMATION:
; APPLICANT: Tam, Robert C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: REGULATION OF CD28 EXPRESSION
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Crockett & Fish
; STREET: 3000 S. Augusta Court
; CITY: La Habra
; STATE: California
; COUNTRY: United States of America
; ZIP: 90631

```

COMPUTER READABLE FORM: 0033  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,878B  
FILING DATE: 13-SEP-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Fish, Robert D.  
REGISTRATION NUMBER: 33,880  
REFERENCE/DOCKET NUMBER: 213/003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-525-3433  
TELEFAX: 714-525-3303  
TELEX:

Query Match 68.0%; Score 20.4; DB 2; Length 1064;  
Best Local Similarity 95.5%; Pred. No. 2.7;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy           9 CTTGTCAGGACGATAGGCTGC 30  
             | | | | | | | | | |  
Db         316 CGTGTCAAGGACGATAGGCTGC 295

### RESULT 3

US-09-023-655-1316/c  
 ; Sequence 1316, Application US/09023655  
 ; Patent No. 6607879  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cocks, Benjamin G.  
 ; APPLICANT: Susan G. Stuart  
 ; APPLICANT: Jeffrey J. Seilhamer  
 ; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
 ; TITLE OF INVENTION: EXPRESSION  
 ; NUMBER OF SEQUENCES: 1508  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 ; STREET: 3174 PORTER DRIVE  
 ; CITY: PALO ALTO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655

Query Match	68.0%	Score 20.4;	DB 4;	Length 1514;
Best Local Similarity	95.5%	Pred. No. 2.9;		
Matches 21; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

**Qy**            9   CTTGT CAGGAGCGATAGGCTGC   30  
               |  
**Db**            766 CGTGT CAGGAGCGATAGGCTGC   745

## RESULT 4

RESOLVED 1  
US-08-549-846-3/c  
; Sequence 3, Application US/08549846  
; Patent No. 6083751  
; GENERAL INFORMATION:  
; APPLICANT: FELDHAUS, ANDREW L.  
; APPLICANT: JONES, LORI A.  
; TITLE OF INVENTION: CHIMERIC RECEPTORS FOR THE GENERATION OF  
; TITLE OF INVENTION: SELECTIVITY-ACTIVATABLE T(h)-INDEPENDENT CYTOTOXIC T CELLS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/549,846  
; FILING DATE: 01-NOV-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DYLAN, TYDER  
; REGISTRATION NUMBER: 37,612  
; REFERENCE/DOCKET NUMBER: 22627-20013.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792



## US-09-716-129-46

Query Match 63.3%; Score 19; DB 4; Length 1189;  
Best Local Similarity 81.5%; Pred. No. 13;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GAGGATCTTGTCCAGGAGCGATAGGCTG 29  
Db 361 GAGGATTTGCCAGGTGGGATGGGCTG 335

## RESULT 8

US-09-774-528-129/c  
; Sequence 129, Application US/09774528  
; Patent No. 6743619  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Zhang, Jie  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6743619el Nucleic Acids and  
; Polypeptides  
; FILE REFERENCE: 802  
; CURRENT APPLICATION NUMBER: US/09/774,528  
; CURRENT FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 129  
; LENGTH: 4758  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (167)..(4381)  
US-09-774-528-129

Query Match 63.3%; Score 19; DB 4; Length 4758;  
Best Local Similarity 81.5%; Pred. No. 18;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AGGATCTTGTCCAGGAGCGATAGGCTGC 30  
Db 897 AGGATCTCTCTCAGGAGCCATGGGCTGC 871

## RESULT 9

US-08-812-946A-7  
; Sequence 7, Application US/08812946A  
; Patent No. 6221637  
; GENERAL INFORMATION:  
; APPLICANT: Tsunekaki HIDA et al.  
; TITLE OF INVENTION: XANTHENE DERIVATIVES, THEIR PRODUCTION AND  
; USE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/812,946A  
; FILING DATE: March 4, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 29 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Other nucleic acid, Synthetic DNA  
US-08-812-946A-7

Query Match 60.0%; Score 18; DB 3; Length 29;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 TCAGGAGCGATAGGCTGC 30  
Db 8 TCAGGAGCGATAGGCTGC 25

## RESULT 10

US-09-270-767-1027  
; Sequence 1027, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1027  
; LENGTH: 268  
; TYPE: DNA  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-1027

Query Match 60.0%; Score 18; DB 4; Length 268;  
Best Local Similarity 80.8%; Pred. No. 26;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 GGATCTTGTCCAGGAGCGATAGGCTGC 30  
Db 114 GGATCTGGTCCAGGAGCGAGCTGC 139

## RESULT 11

US-09-270-767-16309  
; Sequence 16309, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17

```
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16309
; LENGTH: 268
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16309

Query Match          60.0%; Score 18; DB 4; Length 268;
Best Local Similarity 80.8%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GGATCTTGTCCAGGAGCGATAGGCTGC 30
Db 114 GGATCTGGTGAGGAGCGACGAGCTGC 139

RESULT 12
US-09-270-767-2215
; Sequence 2215, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2215
; LENGTH: 746
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-2215

Query Match          60.0%; Score 18; DB 4; Length 746;
Best Local Similarity 80.8%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GGATCTTGTCCAGGAGCGATAGGCTGC 30
Db 114 GGATCTGGTGAGGAGCGACGAGCTGC 139

RESULT 13
US-09-270-767-17497
; Sequence 17497, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17497
; LENGTH: 746
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-17497

Query Match          60.0%; Score 18; DB 4; Length 746;
Best Local Similarity 80.8%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GGATCTTGTCCAGGAGCGATAGGCTGC 30
Db 114 GGATCTGGTGAGGAGCGACGAGCTGC 139

RESULT 14
US-09-270-767-15260/c
```

```
; Sequence 15260, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15260
; LENGTH: 1498
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15260

Query Match          60.0%; Score 18; DB 4; Length 1498;
Best Local Similarity 80.8%; Pred. No. 40;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GGATCTTGTCCAGGAGCGATAGGCTGC 30
Db 502 GGATCTTGTGACGATCGATTGGCGGC 477

RESULT 15
US-09-181-339-4
; Sequence 4, Application US/09181339
; Patent No. 6610827
; GENERAL INFORMATION:
; APPLICANT: Forsayeth, John R.
; APPLICANT: Zhao, Byron
; APPLICANT: Chavez, Raymond C.
; TITLE OF INVENTION: POTASSIUM CHANNEL SUBUNIT POLYPEPTIDE
; TITLE OF INVENTION: AND POLYNUCLEOTIDE COMPOSITIONS AND USES THEREFOR
; FILE REFERENCE: 5865-0033.30
; CURRENT APPLICATION NUMBER: US/09/181,339
; CURRENT FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: US 60/063,450
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 774
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(772)
US-09-181-339-4

Query Match          59.3%; Score 17.8; DB 4; Length 774;
Best Local Similarity 75.9%; Pred. No. 43;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTCCAGGAGCGATAGGCTGC 30
Db 743 CGAGAAACCTCTCAGGAGGCGACAGTCTGC 771

RESULT 16
US-09-221-017B-726/c
; Sequence 726, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
```

COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221.017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 726:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1974 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...1974  
US-09-221-017B-726

Query Match 59.3%; Score 17.8; DB 4; Length 1974;  
Best Local Similarity 75.9%; Pred. No. 54;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTCCAGGAGCGATAGGCTG 29  
Db 983 TCGGGAGCTTGTCCGAGGATACGCCG 955

## RESULT 17

US-09-252-991A-7065/c  
Sequence 7065, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 7065

LENGTH: 1758  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7065

Query Match 58.0%; Score 17.4; DB 4; Length 1758;  
Best Local Similarity 77.8%; Pred. No. 81;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTCCAGGAGCGATAGGCTG 29  
Db 125 GAGGATCCCGTCATGACCGATACCTG 99

## RESULT 18

US-09-252-991A-7009  
Sequence 7009, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 7009  
LENGTH: 2910  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7009

Query Match 58.0%; Score 17.4; DB 4; Length 2910;  
Best Local Similarity 77.8%; Pred. No. 92;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTCCAGGAGCGATAGGCTG 29  
Db 19 GAGGATCCCGTCATGACCGATACCTG 45

## RESULT 19

US-09-252-991A-7120/c  
Sequence 7120, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 7120  
LENGTH: 3186  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7120

Query Match 58.0%; Score 17.4; DB 4; Length 3186;  
Best Local Similarity 77.8%; Pred. No. 94;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTCCAGGAGCGATAGGCTG 29  
Db 125 GAGGATCCCGTCATGACCGATACCTG 99





; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 1492  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-741-238-11

Query Match 57.3%; Score 17.2; DB 4; Length 1492;  
Best Local Similarity 73.3%; Pred. No. 97;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
| | | | | | | | | | | | | | | | | | | | | |  
Db 725 TAGGGTCCCTGTTCAGGGCGGTACGCTGC 696

RESULT 25  
US-09-600-991-3/c  
; Sequence 3, Application US/09600991  
; Patent No. 6551991  
; GENERAL INFORMATION:  
; APPLICANT: MEDICO, Enzo  
; APPLICANT: MICHELI, Paolo  
; APPLICANT: COLLESI, Chiara  
; APPLICANT: CASELLI, Gianfranco  
; APPLICANT: COMOGGIO, Paolo  
; TITLE OF INVENTION: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP  
; FILE REFERENCE: 0471-0162P  
; CURRENT APPLICATION NUMBER: US/09/600,991  
; CURRENT FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 1692  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Metron F-1 DNA coding sequence  
US-09-600-991-3

Query Match 57.3%; Score 17.2; DB 4; Length 1692;  
Best Local Similarity 73.3%; Pred. No. 1e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
| | | | | | | | | | | | | | | | | | | | | |  
Db 695 TCGAGGAACCTTGCCCGGCTCGAAGGGGTGC 666

RESULT 26  
US-09-600-991-21/c  
; Sequence 21, Application US/09600991  
; Patent No. 6551991  
; GENERAL INFORMATION:  
; APPLICANT: MEDICO, Enzo  
; APPLICANT: MICHELI, Paolo  
; APPLICANT: COLLESI, Chiara  
; APPLICANT: CASELLI, Gianfranco  
; APPLICANT: COMOGGIO, Paolo  
; TITLE OF INVENTION: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP  
; FILE REFERENCE: 0471-0162P  
; CURRENT APPLICATION NUMBER: US/09/600,991  
; CURRENT FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 21  
; LENGTH: 1709  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Metron F-1  
US-09-600-991-21

Query Match 57.3%; Score 17.2; DB 4; Length 1709;  
Best Local Similarity 73.3%; Pred. No. 1e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
| | | | | | | | | | | | | | | | | | | | | |  
Db 705 TCGAGGAACCTTGCCCGGCTCGAAGGGGTGC 676

RESULT 27  
US-07-882-925A-3/c  
; Sequence 3, Application US/07882925A  
; Patent No. 5315000  
; GENERAL INFORMATION:  
; APPLICANT: Degen, Sandra J. F.  
; TITLE OF INVENTION: Gene for a growth factor and its cDNA and  
; TITLE OF INVENTION: protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gregory Lunn  
; STREET: Wood, Herron & Evans, 2700 Carew Tower  
; CITY: Cincinnati  
; STATE: Ohio  
; COUNTRY: USA  
; ZIP: 45202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 6.0.3  
; SOFTWARE: Microsoft Word 4.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/882,925A  
; FILING DATE: 19920514  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lunn, Gregory  
; REGISTRATION NUMBER: 29,945  
; REFERENCE/DOCKET NUMBER: CMC 57  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (513) 241-2324  
; TELEFAX: (513) 421-7269  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2021 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; ANTI-SENSE: no  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; DEVELOPMENTAL STAGE: fetal  
; TISSUE TYPE: liver  
; IMMEDIATE SOURCE:  
; LIBRARY: cDNA  
; CLONE: #19  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: human 3p21/D3F15S2  
; FEATURE:  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: This sequence is a variant where two regions  
; OTHER INFORMATION: were found to be deleted when compared to SEQ ID NO:1.  
; PUBLICATION INFORMATION:  
; RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 2021  
US-07-882-925A-3

Query Match 57.3%; Score 17.2; DB 1; Length 2021;  
Best Local Similarity 73.3%; Pred. No. 1e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
| | | | | | | | | | | | | | | | | | | | | |  
Db 660 TCGAGGAACCTTGCCCGGCTCGAAGGGGTGC 631

## RESULT 28

US-08-184-012C-3/c

; Sequence 3, Application US/08184012C

; Patent No. 5606029

; GENERAL INFORMATION:

; APPLICANT: Degen, Sandra J. F.

; TITLE OF INVENTION: Gene for a growth factor and its cDNA and

; TITLE OF INVENTION: protein

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Gregory Lunn

; STREET: Wood, Herron &amp; Evans, 2700 Carew Tower

; CITY: Cincinnati

; STATE: Ohio

; COUNTRY: USA

; ZIP: 45202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: Macintosh 7.5.2

; SOFTWARE: Microsoft Word 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/184,012C

; FILING DATE: 1/18/94

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Lunn, Gregory

; REGISTRATION NUMBER: 29,945

; REFERENCE/DOCKET NUMBER: CMC 57

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (513) 241-2324

; TELEFAX: (513) 421-7269

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2021 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ANTI-SENSE: no

; ORIGINAL SOURCE:

; ORGANISM: human

; DEVELOPMENTAL STAGE: fetal

; TISSUE TYPE: liver

; IMMEDIATE SOURCE:

; LIBRARY: cDNA

; CLONE: #19

; POSITION IN GENOME:

; CHROMOSOME/SEGMENT: human 3p21/D3F15S2

; FEATURE:

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: This sequence is a variant where two regions

; OTHER INFORMATION: were found to be deleted when compared to SEQ ID NO:1.

; PUBLICATION INFORMATION:

; RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 2021

US-08-184-012C-3

Query Match

Best Local Similarity 57.3%; Score 17.2; DB 1; Length 2021;

Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY

1 TCGAGGATCTTGTTCAGGAGCGGATAGGCTGC 30

Db

660 TCGAGGACTTGTCCCGCTCGAAGGGGTGC 631

## RESULT 29

US-09-600-991-19/c

; Sequence 19, Application US/09600991

; Patent No. 6551991

; GENERAL INFORMATION:

; APPLICANT: MEDICO, Enzo

; APPLICANT: MICHELLE, Paolo

; APPLICANT: COLLESI, Chiara

; APPLICANT: CASELLI, Gianfranco

; APPLICANT: COMOGGIO, Paolo

; TITLE OF INVENTION: RECOMBINANT PROTEINS DERIVED FROM HGF AND MSP

; FILE REFERENCE: 0471-0162P

; CURRENT APPLICATION NUMBER: US/09/600,991

; CURRENT FILING DATE: 2001-08-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 19

; LENGTH: 2136

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: CDS

; LOCATION: (1)..(2136)

; OTHER INFORMATION:

; US-09-600-991-19

Query Match 57.3%; Score 17.2; DB 4; Length 2136;

Best Local Similarity 73.3%; Pred. No. 1.1e+02;

Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY

1 TCGAGGATCTTGTTCAGGAGCGGATAGGCTGC 30

Db

695 TCGAGGACTTGTCCCGCTCGAAGGGGTGC 666

## RESULT 30

US-09-601-040A-11/c

; Sequence 11, Application US/09601040A

; Patent No. 6730657

; GENERAL INFORMATION:

; APPLICANT: MEDICO, Enzo et al.

; TITLE OF INVENTION: Recombinant Proteins from HGF and MSP

; FILE REFERENCE: 0471-0161P

; CURRENT APPLICATION NUMBER: US/09/601,040A

; CURRENT FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 11

; LENGTH: 2136

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-601-040A-11

Query Match

Best Local Similarity 57.3%; Score 17.2; DB 4; Length 2136;

Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY

1 TCGAGGATCTTGTTCAGGAGCGGATAGGCTGC 30

Db

695 TCGAGGACTTGTCCCGCTCGAAGGGGTGC 666

## RESULT 31

US-07-882-925A-4/c

; Sequence 4, Application US/07882925A

; Patent No. 5315000

; GENERAL INFORMATION:

; APPLICANT: Degen, Sandra J. F.

; TITLE OF INVENTION: Gene for a growth factor and its cDNA and

; TITLE OF INVENTION: protein

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Gregory Lunn

; STREET: Wood, Herron &amp; Evans, 2700 Carew Tower

; CITY: Cincinnati

; STATE: Ohio

; COUNTRY: USA

; ZIP: 45202

FILING DATE: 1/18/94  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Gregory  
REGISTRATION NUMBER: 29,945  
REFERENCE/DOCKET NUMBER: CMC 57  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (513) 241-2324  
TELEFAX: (513) 421-7269  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2188 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: mouse  
STRAIN: C57BL/6  
DEVELOPMENTAL STAGE: adult  
TISSUE TYPE: liver  
IMMEDIATE SOURCE:  
LIBRARY: cDNA  
CLONE: ML5-2  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: mouse 9, Hgf1 locus  
MAP POSITION: 11f-Gnal-2-Hgf1-Cck  
FEATURE:  
IDENTIFICATION METHOD: experimental  
PUBLICATION INFORMATION:  
RELEVANT RESIDUES IN SEQ ID NO: 4: 1 TO 2188  
US-08-184-012C-4

Query Match 57.3%; Score 17.2; DB 1; Length 2188;  
Best Local Similarity 73.3%; Pred. No. 1.1e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0;

QY 1 TCGAGGATCTTGTCCAGGAGCGATAGGCTGC 30  
Db 651 TCTAGGACCTTTCAGGCTGCGAAGGGTGC 622

RESULT 33  
US-08-666-082B-2/c  
Sequence 2, Application US/08666082B  
Patent No. 5916770  
GENERAL INFORMATION:  
APPLICANT: YOSHIKAWA, WATARU  
APPLICANT: SHIMONISHI, MANABU  
APPLICANT: IWAMOTO, JUNKO  
APPLICANT: TAKEHARA, TOYOHIRO  
APPLICANT: HAGIYA, MICHIO  
TITLE OF INVENTION: MACROPHAGE STIMULATING PROTEIN VARIANT AND  
TITLE OF INVENTION: METHOD FOR PRODUCING THE SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
STREET: TWO PRUDENTIAL PLAZA, SUITE 4900  
CITY: CHICAGO  
STATE: IL  
COUNTRY: US  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/666,082B  
FILING DATE: 19-JUN-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 7-153309  
FILING DATE: 20-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBERT F. GREEN  
REGISTRATION NUMBER: 27555  
REFERENCE/DOCKET NUMBER: 73843  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2216 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Human  
FEATURE:  
NAME/KEY: sig peptide  
LOCATION: 1..54  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2133  
US-08-666-082B-2

Query Match 57.3%; Score 17.2; DB 2; Length 2216;  
Best Local Similarity 73.3%; Pred. No. 1.1e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGACTTGTTCAGGAGCGATAGGCTGC 30  
||||| ||||| ||||| ||||| |||||  
Db 695 TCGAGGACTTGTCCCGCTCGAGGGGTGC 666

RESULT 34  
US-07-882-925A-1/c  
Sequence 1, Application US/07882925A  
Patent No. 5315000  
GENERAL INFORMATION:  
APPLICANT: Degen, Sandra J. F.  
TITLE OF INVENTION: Gene for a growth factor and its cDNA and  
TITLE OF INVENTION: protein  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gregory Lunn  
STREET: Wood, Herron & Evans, 2700 Carew Tower  
CITY: Cincinnati  
STATE: Ohio  
COUNTRY: USA  
ZIP: 45202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.3  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/882,925A  
FILING DATE: 19920514  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Gregory  
REGISTRATION NUMBER: 29,945  
REFERENCE/DOCKET NUMBER: CMC 57  
TELEPHONE: (513) 241-2324  
TELEFAX: (513) 421-7269  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2219 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:

ORGANISM: human  
DEVELOPMENTAL STAGE: fetal  
TISSUE TYPE: liver  
IMMEDIATE SOURCE:  
LIBRARY: cDNA  
CLONE: #33  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: human 3p21/D3F15S2  
FEATURE:  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: Includes five polymorphisms at the nucleotide  
level; one of which results in an amino acid substitution (nu  
OTHER INFORMATION: 619). Sequence ID NO:2: contains the identical sequence with  
OTHER INFORMATION: polymorphic amino acid.  
PUBLICATION INFORMATION:  
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2219  
US-07-882-925A-1

Query Match 57.3%; Score 17.2; DB 1; Length 2219;  
Best Local Similarity 73.3%; Pred. No. 1.1e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGACTTGTTCAGGAGCGATAGGCTGC 30  
||||| ||||| ||||| ||||| |||||  
Db 679 TCGAGGACTTGTCCCGCTCGAGGGGTGC 650

RESULT 35  
US-07-882-925A-2/c  
Sequence 2, Application US/07882925A  
Patent No. 5315000  
GENERAL INFORMATION:  
APPLICANT: Degen, Sandra J. F.  
TITLE OF INVENTION: Gene for a growth factor and its cDNA and  
TITLE OF INVENTION: protein  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gregory Lunn  
STREET: Wood, Herron & Evans, 2700 Carew Tower  
CITY: Cincinnati  
STATE: Ohio  
COUNTRY: USA  
ZIP: 45202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.3  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/882,925A  
FILING DATE: 19920514  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Gregory  
REGISTRATION NUMBER: 29,945  
REFERENCE/DOCKET NUMBER: CMC 57  
TELEPHONE: (513) 241-2324  
TELEFAX: (513) 421-7269  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2219 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: human  
DEVELOPMENTAL STAGE: fetal  
TISSUE TYPE: liver  
IMMEDIATE SOURCE:  
LIBRARY: cDNA

```
; CLONE: #33
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: human 3p21/D3F15S2
; FEATURE:
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: Includes five polymorphisms at the nucleotide
; OTHER INFORMATION: level; one of which results in an amino acid substitution (nu
; OTHER INFORMATION: 619). Sequence ID NO:1: contains the identical sequence with
; OTHER INFORMATION: polymorphic amino acid.
; PUBLICATION INFORMATION:
; RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2219
US-07-882-925A-2
Query Match 57.3%; Score 17.2; DB 1; Length 2219;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 TCGAGGATCTTGTCCAGGCGGATAGGCTGC 30
Db 679 TCGAGGAACCTTCCCGGCTCGAAGGGGTGC 650
RESULT 36
US-08-184-012C-1/c
; Sequence 1, Application US/08184012C
; Patent No. 5606029
; GENERAL INFORMATION:
; APPLICANT: Degen, Sandra J. F.
; TITLE OF INVENTION: Gene for a growth factor and its cDNA and
; TITLE OF INVENTION: protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gregory Lunn
; STREET: Wood, Herron & Evans, 2700 Carew Tower
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5.2
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,012C
; FILING DATE: 1/18/94
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Gregory
; REGISTRATION NUMBER: 29,945
; REFERENCE/DOCKET NUMBER: CMC 57
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 241-2324
; TELEFAX: (513) 421-7269
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; DEVELOPMENTAL STAGE: fetal
; TISSUE TYPE: liver
; IMMEDIATE SOURCE:
; LIBRARY: cDNA
; CLONE: #33
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: human 3p21/D3F15S2
; FEATURE:
; IDENTIFICATION METHOD: experimental
```

```
; OTHER INFORMATION: Includes five polymorphisms at the nucleotide
; OTHER INFORMATION: level; one of which results in an amino acid substitution (nu
; OTHER INFORMATION: 619). Sequence ID NO:2: contains the identical sequence with
; Patent No. 5606029
; OTHER INFORMATION: polymorphic amino acid.
; PUBLICATION INFORMATION:
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2219
US-08-184-012C-1
Query Match 57.3%; Score 17.2; DB 1; Length 2219;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 TCGAGGATCTTGTCCAGGCGGATAGGCTGC 30
Db 679 TCGAGGAACCTTCCCGGCTCGAAGGGGTGC 650
RESULT 37
US-08-184-012C-2/c
; Sequence 2, Application US/08184012C
; Patent No. 5606029
; GENERAL INFORMATION:
; APPLICANT: Degen, Sandra J. F.
; TITLE OF INVENTION: Gene for a growth factor and its cDNA and
; TITLE OF INVENTION: protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gregory Lunn
; STREET: Wood, Herron & Evans, 2700 Carew Tower
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5.2
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,012C
; FILING DATE: 1/18/94
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Gregory
; REGISTRATION NUMBER: 29,945
; REFERENCE/DOCKET NUMBER: CMC 57
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 241-2324
; TELEFAX: (513) 421-7269
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; DEVELOPMENTAL STAGE: fetal
; TISSUE TYPE: liver
; IMMEDIATE SOURCE:
; LIBRARY: cDNA
; CLONE: #33
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: human 3p21/D3F15S2
; FEATURE:
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: Includes five polymorphisms at the nucleotide
; OTHER INFORMATION: level; one of which results in an amino acid substitution (nu
; OTHER INFORMATION: 619). Sequence ID NO:1: contains the identical sequence with
; Patent No. 5606029
```

```
;
; OTHER INFORMATION: polymorphic amino acid.
; PUBLICATION INFORMATION:
; RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2219
US-08-184-012C-2

Query Match 57.3%; Score 17.2; DB 1; Length 2219;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
Db 679 TCGAGGAACTTGCCTCGCTCGAAGGGGTGC 650

RESULT 38
US-08-334-177-1/c
; Sequence 1, Application US/08334177
; Patent No. 5696086
; GENERAL INFORMATION:
; APPLICANT: Avraham, Hava Karsenty
; APPLICANT: Godowski, Paul J.
; TITLE OF INVENTION: Methods and Kits Using Macrophage Stimulating Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/334,177
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: 912
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5416
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2232 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-334-177-1

Query Match 57.3%; Score 17.2; DB 1; Length 2232;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
Db 705 TCGAGGAACTTGCCTCGCTCGAAGGGGTGC 676

RESULT 39
PCT-US95-13830-1/c
; Sequence 1, Application PC/TUS9513830
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: New England Deaconess Hospital Corp.
```

```
;
; TITLE OF INVENTION: Methods and Kits Using Macrophage Stimulating
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13830
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: P0912PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5416
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2232 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; PCT-US95-13830-1

Query Match 57.3%; Score 17.2; DB 5; Length 2232;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
Db 705 TCGAGGAACTTGCCTCGCTCGAAGGGGTGC 676

RESULT 40
US-07-882-925A-7/c
; Sequence 7, Application US/07882925A
; Patent No. 5315000
; GENERAL INFORMATION:
; APPLICANT: Degen, Sandra J. F.
; TITLE OF INVENTION: Gene for a growth factor and its cDNA and
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gregory Lunn
; STREET: Wood, Herron & Evans, 2700 Carew Tower
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.3
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/882,925A
; FILING DATE: 19920514
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Gregory
; REGISTRATION NUMBER: 29,945
; REFERENCE/DOCKET NUMBER: CMC 57
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 241-2324
; TELEFAX: (513) 421-7269
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2262 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; DESCRIPTION: Identical to sequence ID NO: 1: with 5' and 3'
; ADAPTERS: adaptors added to make a full-length CDNA
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; DEVELOPMENTAL STAGE: fetal
; TISSUE TYPE: liver
; IMMEDIATE SOURCE:
; LIBRARY: CDNA
; CLONE: #33 including 5' and 3' adaptors
; PUBLICATION INFORMATION:
; RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 2262
;
US-07-882-925A-7

```

```

Query Match      57.3%; Score 17.2; DB 1; Length 2262;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

Qy      1 TCGAGGATCTTTCAGGAGCGATAGGCTGC 30
      ||||| ||||| ||||| ||||| |||||
Db     710 TCGAGGAACTTCCCGGCTCGAAGGGGTGC 681

```

```

Search completed: January 7, 2005, 11:00:37
Job time : 56.2105 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 11:00:44 ; Search time 1275.26 Seconds  
(without alignments)  
134.979 Million cell updates/sec

Title: US-09-786-502A-8  
Perfect score: 30  
Sequence: 1 tcgagagattcttcaggagcgtaggctgc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4293498 seqs, 2868903791 residues

Total number of hits satisfying chosen parameters: 8586996

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*  
1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/prodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/prodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
19: /cgn2\_6/prodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
20: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
21: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	US-08-940-544-2	Sequence 2, Appli
2	20.4	68.0	1514	US-09-738-546-1	Sequence 1, Appli
3	20.4	68.0	1514	US-09-836-544-7	Sequence 7, Appli
4	20.4	68.0	1514	US-10-207-655-98	Sequence 98, Appl
5	20.4	68.0	1514	US-10-641-643-1316	Sequence 1316, Ap
6	20.4	68.0	3803	US-10-076-934-1	Sequence 1, Appli
7	20.4	68.0	3804	US-09-997-722-257	Sequence 257, App
8	20.4	68.0	3804	US-10-143-238-1	Sequence 1, Appli
9	20.4	68.0	3806	US-09-835-297-3	Sequence 3, Appli
10	20.4	68.0	51365	US-09-997-722-256	Sequence 256, App
11	20.2	67.3	36	US-10-448-256-8	Sequence 8, Appli
12	20	66.7	32	US-10-273-762-2	Sequence 2, Appli

c 13	19.4	64.7	1329	17	US-10-437-963-23123	Sequence 23123, A
c 14	19.4	64.7	4467	17	US-10-437-963-23127	Sequence 23127, A
c 15	19.4	64.7	4509	17	US-10-437-963-28652	Sequence 28652, A
c 16	19.4	64.7	430442	18	US-10-417-375-128	Sequence 128, App
c 17	19	63.3	866	10	US-09-986-480-134	Sequence 134, App
c 18	19	63.3	1189	16	US-10-632-983-46	Sequence 46, Appl
c 19	19	63.3	4758	15	US-10-120-988-129	Sequence 129, App
c 20	19	63.3	303172	18	US-10-719-993-6890	Sequence 6890, Ap
c 21	18.8	62.7	392	16	US-10-424-599-90830	Sequence 90830, A
c 22	18.8	62.7	542	16	US-10-260-238-441	Sequence 441, App
c 23	18.8	62.7	609	15	US-10-369-493-38650	Sequence 38650, A
c 24	18.8	62.7	630	15	US-10-369-493-35152	Sequence 35152, A
c 25	18.8	62.7	630	15	US-10-369-493-38079	Sequence 38079, A
c 26	18.8	62.7	630	15	US-10-369-493-38204	Sequence 38204, A
c 27	18.8	62.7	815	17	US-10-437-963-48174	Sequence 48174, A
c 28	18.8	62.7	1431	15	US-10-369-493-44430	Sequence 44430, A
c 29	18.8	62.7	2115	17	US-10-437-963-67598	Sequence 67598, A
c 30	18.4	61.3	48	15	US-10-273-762-7	Sequence 7, Appli
c 31	18.4	61.3	2609	15	US-10-172-118-212	Sequence 212, App
c 32	18.4	61.3	2809	16	US-10-342-887-212	Sequence 212, App
c 33	18.4	61.3	9025608	15	US-10-156-761-1	Sequence 1, Appli
c 34	18.2	60.7	667	18	US-10-425-115-34870	Sequence 34870, A
c 35	18.2	60.7	1523	16	US-10-425-114-22914	Sequence 22914, A
c 36	18.2	60.7	1604	18	US-10-425-115-34867	Sequence 34867, A
c 37	18.2	60.7	52679	13	US-10-087-192-553	Sequence 553, App
c 38	18	60.0	328	16	US-10-448-256-6	Sequence 6, Appli
c 39	18	60.0	388	9	US-09-960-352-10773	Sequence 10773, A
c 40	18	60.0	597	13	US-10-027-632-107348	Sequence 107348, A
c 41	18	60.0	597	15	US-10-027-632-107348	Sequence 107348, A
c 42	18	60.0	663	11	US-09-997-722-258	Sequence 258, App
c 43	18	60.0	727	13	US-10-027-632-27997	Sequence 27997, A
c 44	18	60.0	727	15	US-10-027-632-27997	Sequence 27997, A
c 45	18	60.0	809	13	US-10-027-632-9959	Sequence 9959, Ap

ALIGNMENTS

RESULT 1

US-08-940-544-2  
; Sequence 2, Application US/08940544B  
; Publication No. US20020018783A1  
; GENERAL INFORMATION:  
; APPLICANT: SADELAIN, MICHEL  
; APPLICANT: CHEUNG, NAI-KONG V.  
; APPLICANT: KRAUSE, ANJA  
; APPLICANT: GUO, HONG-FEN  
; TITLE OF INVENTION: FUSION PROTEINS OF A SINGLE CHAIN ANTIBODY AND CD28 AND  
; TITLE OF INVENTION: FUSION THEREOF  
; FILE REFERENCE: MSK.P-035-US  
; CURRENT APPLICATION NUMBER: US/08/940,544B  
; CURRENT FILING DATE: 1997-09-30  
; EARLIER APPLICATION NUMBER: PCT/US97/04427  
; EARLIER FILING DATE: 1997-03-20  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 30  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; OTHER INFORMATION: Downstream primer for PCR amplification.  
US-08-940-544-2

Query Match 100.0%; Score 30; DB 8; Length 30;  
Best Local Similarity 100.0%; Pred. No. 0.00077; Indels 0; Gaps 0;  
Matches 30; Conservative 0; Mismatches 0;  
QY 1 TCGAGGATCTTCTCAGGAGCGTAGGCTGC 30  
DB 1 TCGAGGATCTTCTCAGGAGCGTAGGCTGC 30

```
RESULT 2
US-09-738-546-1/c
; Sequence 1, Application US/09738546
; Patent No. US20020006403A1
; GENERAL INFORMATION:
; APPLICANT: YU, XUE-ZHONG
; APPLICANT: ANASTETI, CIAUDIO
; TITLE OF INVENTION: CD28-SPECIFIC ANTIBODY COMPOSITIONS FOR USE IN METHODS
; FILE REFERENCE: PHC0.007US
; CURRENT APPLICATION NUMBER: US/09/738,546
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/170,857.
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-738-546-1

Query Match      68.0%; Score 20.4; DB 9; Length 1514;
Best Local Similarity 95.5%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      9 CTTGTCAGGAGCGATAGGCTGC 30
DB      766 CGTGTACGAGCGATAGGCTGC 745

RESULT 3
US-09-836-544-7/c
; Sequence 7, Application US/09836544
; Publication No. US20040072283A1
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: Rapid Immunoselection Cloning Method
; FILE REFERENCE: 11-88L
; CURRENT APPLICATION NUMBER: US/09/836,544
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 07/983,647
; PRIOR FILING DATE: 1992-12-01
; PRIOR APPLICATION NUMBER: US 07/553,759
; PRIOR FILING DATE: 1990-07-13
; PRIOR APPLICATION NUMBER: US 07/498,809
; PRIOR FILING DATE: 1990-03-23
; PRIOR APPLICATION NUMBER: US 07/379,076
; PRIOR FILING DATE: 1989-07-13
; PRIOR APPLICATION NUMBER: US 07/160,416
; PRIOR FILING DATE: 1988-02-25
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (100)..(759)
US-09-836-544-7

Query Match      68.0%; Score 20.4; DB 11; Length 1514;
Best Local Similarity 95.5%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      9 CTTGTCAGGAGCGATAGGCTGC 30
DB      766 CGTGTACGAGCGATAGGCTGC 745

RESULT 4
US-10-207-655-98/c
; Sequence 98, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden-Ledbetter, Martha S.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207,655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 98
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-207-655-98

Query Match      68.0%; Score 20.4; DB 15; Length 1514;
Best Local Similarity 95.5%; Pred. No. 23;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      9 CTTGTCAGGAGCGATAGGCTGC 30
DB      766 CGTGTACGAGCGATAGGCTGC 745

RESULT 5
US-10-641-643-1316/c
; Sequence 1316, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/641,643
; FILING DATE: 14-Aug-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1316:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g338444
; SEQUENCE DESCRIPTION: SEQ ID NO: 1316 :
```

Qy 9 CTTGTcAGGAGCGATAGGCTGC 30  
| | | | |  
pb 889 CGTGTcAGGAGCGATAGGCTGC 868

US-09-997-722-256/c  
; Sequence 256, Application US/09997722  
; Publication No. US20040072154A1  
; GENERAL INFORMATION:

; APPLICANT: Morris, David  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71171/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/09/997,722  
; CURRENT FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 301  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 256  
; LENGTH: 51365  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-997-722-256

Query Match 68.0%; Score 20.4; DB 11; Length 51365;  
Best Local Similarity 95.5%; Pred. No. 26;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 CTTGTGAGGAGCGATAGGCTGC 30  
DB 38448 CGTGCAGGAGCGATAGGCTGC 38427

RESULT 11  
US-10-448-256-8  
; Sequence 8, Application US/10448256  
; Publication No. US20040043401A1  
; GENERAL INFORMATION:  
; APPLICANT: Sadelain, Michel  
; APPLICANT: Brentjens, Renier  
; APPLICANT: Maher, John  
; TITLE OF INVENTION: Chimeric T Cell Receptors  
; FILE REFERENCE: MSK.P-058  
; CURRENT APPLICATION NUMBER: US/10/448,256  
; CURRENT FILING DATE: 2003-05-28  
; PRIOR APPLICATION NUMBER: 60/383,872  
; PRIOR FILING DATE: 2002-05-28  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 8  
; LENGTH: 36  
; TYPE: DNA  
; ORGANISM: human  
US-10-448-256-8

Query Match 67.3%; Score 20.2; DB 16; Length 36;  
Best Local Similarity 88.0%; Pred. No. 24;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GATCTGTGAGGAGCGATAGGCTGC 30  
DB 2 GCTCGAGTCAAGGAGCGATAGGCTGC 26

RESULT 12  
US-10-273-762-2  
; Sequence 2, Application US/10273762  
; Publication No. US20030147881A1  
; GENERAL INFORMATION:  
; APPLICANT: CHEUNG, Nai-Kong V  
; APPLICANT: GUO, Hong-fen  
; TITLE OF INVENTION: METHOD FOR PREPARATION OF SINGLE CHAIN ANTIBODIES  
; FILE REFERENCE: 676-A-PCT  
; CURRENT APPLICATION NUMBER: US/10/273,762  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: 60/330,396  
; PRIOR FILING DATE: 2001-10-17  
; PRIOR APPLICATION NUMBER: 10/097,558  
; PRIOR FILING DATE: 2002-03-08

; PRIOR APPLICATION NUMBER: PCT/US01/32565  
; PRIOR FILING DATE: 2001-10-18  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 32  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-273-762-2

Query Match 66.7%; Score 20; DB 15; Length 32;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TGTGAGGAGCGATAGGCTGC 30  
DB 13 TGTGAGGAGCGATAGGCTGC 32

RESULT 13  
US-10-437-963-23123/c  
; Sequence 23123, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 23123  
; LENGTH: 1329  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_28233C.1  
US-10-437-963-23123

Query Match 64.7%; Score 19.4; DB 17; Length 1329;  
Best Local Similarity 79.3%; Pred. No. 64;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTGAGGAGCGATAGGCTGC 30  
DB 651 CCAGGATCTCTCCAGGAGCGAGAGCTGC 623

RESULT 14  
US-10-437-963-23127  
; Sequence 23127, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14



```

; PRIOR FILING DATE: 1998-02-26
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: Patentin version 2
; SEQ ID NO 46
; LENGTH: 1189
; TYPE: DNA
; ORGANISM: Homo.sapiens
US-10-632-983-46

```

Query Match 63.3%; Score 19; DB 16; Length 1189;  
Best Local Similarity 81.5%; Pred. No. 97;  
Matches 22; Conservative 0; Mismatches 5; Indels

Qy 3 GAGGATCTTGT CAGGAGCGATAGGCTG 29  
361 GAGGATTTTCCAGGTGGGATGGGCTG 335

Db

RESULT 19  
US-10-120-988-129/c  
; Sequence 129, Application US/10120988  
; Publication No. US20030219745A1

/ APPLICANT: Tang, Y. Tom  
 / APPLICANT: Goodrich, Ryle  
 / APPLICANT: Liu, Chenghua  
 / APPLICANT: Ren, Feiyun  
 / APPLICANT: Wang, Dunrui  
 / APPLICANT: Dmanac, Radoje T.  
 / TITLE OF INVENTION: No. US20030219745A1ele Nucleic Acids and  
 / TITLE OF INVENTION: Polypeptides  
 / FILE REFERENCE: 802CON  
 / CURRENT APPLICATION NUMBER: US/10/120,988  
 / CURRENT FILING DATE: 2002-04-11  
 / PRIOR APPLICATION NUMBER: 09/774,528  
 / PRIOR FILING DATE: 2001-01-30  
 / NUMBER OF SEQ ID NOS: 441

```

; NAME/KEY: CDS
; LOCATION: (167)..(4381)
US-10-120-988-129

```

Query Match 63.3%; Score 19; DB 15; Length 4758;  
Best Local Similarity 81.5%; Pred. No. 1e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels

QY 4 AGGATCTTGTCTCAGGAGCGATAGGCTGC 30  
|||||  
db 897 AGGATCTCTCTCAGGAGCCATGGCCTGC 871

RESULT 20  
US-10-719-993-6890/c  
; Sequence 6890, Application US/10719993  
; Publication No. US20040265849A1

```

/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001496
/ CURRENT APPLICATION NUMBER: US/10/719,993
/ CURRENT FILING DATE: 2003-11-24
/ NUMBER OF SEQ ID NOS: 5342
/ SOFTWARE: PASCSEQ for Windows Version 4.0
/ SEQ ID NO 6890
/ LENGTH: 303172
/ TYPE: DNA
/ ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(30317)
; OTHER INFORMATION: n =
US-10-719-993-6890

```

Query Match	63.3%	Score 19;	DB 18;	Length 303172;
Best Local Similarity	81.5%;	Pred. No. 1.2e+02;		
Matches 22;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps	0;			

Qy 1 TCGAGGATCTTGT CAGGAGCGATAGGC 27  
 299749 TCGAGGATCTTGT CAGGAGCGATAGGC 299773  
 Db

RESULT 21  
US-10-424-599-90830  
; Sequence 90830, Application US/10424599  
; Publication No. US20040031072A1

```

/ GENERATED INFORMATION.
/ APPLICANT: La Rosa, Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 90830
/ LENGTH: 392
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_53030C.1
US-10-424-599-90830

```

Query Match	62.7%	Score 18.8;	DB 16;	Length 392;
Best Local Similarity	76.7%	Pred. No. 1.1e+02;		
Matches 23;	Conservative	0;	Mismatches 7;	Indels 0;
Gaps	0;			

Qy 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
|||  
Db 48 TCCAGGATCTGGTCACGAGTGGCAGGCTCC 77

RESULT 22  
US-10-260-238-441/c  
; Sequence 441, Application US/10260238  
; Publication No. US20040016025A1

```

? GENERAL INFORMATION:
? APPLICANT: Budworth, Paul R.
? APPLICANT: Moughamer, Todd G.
? APPLICANT: Briggs, Steven P.
? APPLICANT: Cooper, Bret
? APPLICANT: Glazebrook, Jane
? APPLICANT: Goff, Stephen A.
? APPLICANT: Katagiri, Fumiyaiki
? APPLICANT: Kreps, Joel
? APPLICANT: Provart, Nicholas
? APPLICANT: Ricke, Darrell
? APPLICANT: Zhu, Tong
? TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
? FILE REFERENCE: 60111-NP
? CURRENT APPLICATION NUMBER: US/10/360,238
? CURRENT FILING DATE: 2002-09-26
? PRIOR APPLICATION NUMBER: US 60/325,448
? PRIOR FILING DATE: 2001-09-26
? PRIOR APPLICATION NUMBER: US 60/325,277
? PRIOR FILING DATE: 2001-09-26
? PRIOR APPLICATION NUMBER: US 60/370,620
? PRIOR FILING DATE: 2002-04-04

```

NUMBER OF SEQ ID NOS: 6077  
; SEQ ID NO 441  
; LENGTH: 542  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
US-10-260-238-441

Query Match 62.7%; Score 18.8; DB 16; Length 542;  
Best Local Similarity 90.9%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTTCAGGAGCGAT 23  
DB 519 CGAGGATCTTGTTCAGGAGCGAT 498

RESULT 23  
US-10-369-493-38650  
; Sequence 38650, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 38650  
; LENGTH: 609  
; TYPE: DNA  
; ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-38650

Query Match 62.7%; Score 18.8; DB 15; Length 609;  
Best Local Similarity 76.7%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TCAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
DB 203 TCATCGATCTTGTTCAGGAGCGATATTTCTGC 232

RESULT 24  
US-10-369-493-35152  
; Sequence 35152, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 35152  
; LENGTH: 630  
; TYPE: DNA  
; ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-35152

Query Match 62.7%; Score 18.8; DB 15; Length 630;  
Best Local Similarity 76.7%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TCAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
DB 224 TCATCGATCTTGTTCAGGAGCGATATTTCTGC 253

RESULT 25  
US-10-369-493-38079  
; Sequence 38079, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 38079  
; LENGTH: 630  
; TYPE: DNA  
; ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-38079

Query Match 62.7%; Score 18.8; DB 15; Length 630;  
Best Local Similarity 76.7%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TCAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
DB 224 TCATCGATCTTGTTCAGGAGCGATATTTCTGC 253

RESULT 26  
US-10-369-493-38204  
; Sequence 38204, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 38204  
; LENGTH: 630  
; TYPE: DNA  
; ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-38204

Query Match 62.7%; Score 18.8; DB 15; Length 630;  
Best Local Similarity 76.7%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TCAGGATCTTGTTCAGGAGCGATAGGCTGC 30  
DB 224 TCATCGATCTTGTTCAGGAGCGATATTTCTGC 253

## RESULT 27

US-10-437-963-48174/c  
; Sequence 48174, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 48174  
; LENGTH: 815  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_50875C.1  
US-10-437-963-48174

Query Match 62.7%; Score 18.8; DB 17; Length 815;  
Best Local Similarity 90.9%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTCCAGGAGCGAT 23  
DB 670 CGACGATCTTGGCAGGAGCGAT 649

## RESULT 28

US-10-369-493-44430/c  
; Sequence 44430, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 44430  
; LENGTH: 1431  
; TYPE: DNA  
; ORGANISM: Rhodospseudomonas palustris  
US-10-369-493-44430

Query Match 62.7%; Score 18.8; DB 15; Length 1431;  
Best Local Similarity 90.9%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTCCAGGAGCGA 22  
DB 638 TCGAGGATCTTGTCCGAGATCGA 617

## RESULT 29

US-10-437-963-67598

; Sequence 67598, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 67598  
; LENGTH: 2115  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_6843C.1  
US-10-437-963-67598

Query Match 62.7%; Score 18.8; DB 17; Length 2115;  
Best Local Similarity 76.7%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTCCAGGAGCGATAGCTGC 30  
DB 1517 TCGAGGATCTTATCAAGACAAATATGGTGC 1546

## RESULT 30

US-10-273-762-7/c  
; Sequence 7, Application US/10273762  
; Publication No. US20030147881A1  
; GENERAL INFORMATION:  
; APPLICANT: GUO, Hong-fen  
; APPLICANT: CHEUNG, Nai-Kong V  
; TITLE OF INVENTION: METHOD FOR PREPARATION OF SINGLE CHAIN ANTIBODIES  
; FILE REFERENCE: 676-A-PCT  
; CURRENT APPLICATION NUMBER: US/10/273,762  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: 60/330,396  
; PRIOR FILING DATE: 2001-10-17  
; PRIOR APPLICATION NUMBER: 10/097,558  
; PRIOR FILING DATE: 2002-03-08  
; PRIOR APPLICATION NUMBER: PCT/US01/32565  
; PRIOR FILING DATE: 2001-10-18  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 48  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-273-762-7

Query Match 61.3%; Score 18.4; DB 15; Length 48;  
Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 TGTCCAGGAGCGATAGCTGC 30  
DB 29 TGCAGGAGCGATAGCTGC 10

## RESULT 31

US-10-172-118-212/c  
; Sequence 212, Application US/10172118  
; Publication No. US20030224374A1  
; GENERAL INFORMATION:



```

; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Marc
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 212
; LENGTH: 2609
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK000345
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-212

Query Match      61.3%; Score 18.4; DB 15; Length 2609;
Best Local Similarity 78.6%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    |||||
Db 1670 GAGCATCTTTTCAGGATGGATGGACTGC 1643

RESULT 32
US-10-342-887-212/c
; Sequence 212, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 212
; LENGTH: 2609
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-212

Query Match      61.3%; Score 18.4; DB 16; Length 2609;
Best Local Similarity 78.6%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    |||||
Db 1670 GAGCATCTTTTCAGGATGGATGGACTGC 1643

RESULT 33
US-10-342-887-212/c
; Sequence 212, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 212
; LENGTH: 2609
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-212

Query Match      61.3%; Score 18.4; DB 16; Length 2609;
Best Local Similarity 78.6%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    |||||
Db 1670 GAGCATCTTTTCAGGATGGATGGACTGC 1643
```

```

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      61.3%; Score 18.4; DB 15; Length 9025608;
Best Local Similarity 78.6%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTTCAGGAGCGATAGGCTG 29
    |||||
Db 797584 CGAGTGTCTGTTCAGGAGCGGTAGTTG 797611

RESULT 34
US-10-425-115-34870
; Sequence 34870, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 34870
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_131804C.1
US-10-425-115-34870

Query Match      60.7%; Score 18.2; DB 18; Length 667;
Best Local Similarity 87.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GGATCTTGTTCAGGAGCGATAGGC 27
    |||||
Db 223 GGATCGTGTTCAGAAAGCGAGAGGC 245

RESULT 35
US-10-425-114-22914
```

```
; Sequence 22914, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22914
; LENGTH: 1523
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3592-068-C8_FLI
US-10-425-114-22914

Query Match          60.7%; Score 18.2; DB 16; Length 1523;
Best Local Similarity 87.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      5 GGATCTTGTCAGGAGCGATAGGC 27
Db      876 GGATCGTGTCAGAGCGAGAGGC 898

RESULT 36
US-10-425-115-34867
; Sequence 34867, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 34867
; LENGTH: 1604
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_131801C.1
US-10-425-115-34867

Query Match          60.7%; Score 18.2; DB 18; Length 1604;
Best Local Similarity 87.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      5 GGATCTTGTCAGGAGCGATAGGC 27
Db      898 GGATCGTGTCAGAGCGAGAGGC 920

RESULT 37
US-10-087-192-553
; Sequence 553, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
```

```
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 553
; LENGTH: 52679
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(52679)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-553

Query Match          60.7%; Score 18.2; DB 13; Length 52679;
Best Local Similarity 87.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 AGGATCTTGTCAGGAGCGATAGG 26
Db      1240 AGGATTTTGCAGGAGCAATAGG 1262

RESULT 38
US-10-448-256-6/c
; Sequence 6, Application US/10448256
; Publication No. US20040043401A1
; GENERAL INFORMATION:
; APPLICANT: Sadelain, Michel
; APPLICANT: Brentjens, Renier
; APPLICANT: Maher, John
; TITLE OF INVENTION: Chimeric T Cell Receptors
; FILE REFERENCE: MSK.P-058
; CURRENT APPLICATION NUMBER: US/10/448,256
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/383,872
; PRIOR FILING DATE: 2002-05-28
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 328
; TYPE: DNA
; ORGANISM: human
US-10-448-256-6

Query Match          60.0%; Score 18; DB 16; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      13 TCAGGAGCGATAGGCTGC 30
Db      328 TCAGGAGCGATAGGCTGC 311

RESULT 39
US-09-960-352-10773
; Sequence 10773, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
```

```
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10773
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 46-LIB34-013-Q1-E1-D6
US-09-960-352-10773

Query Match      60.0%; Score 18; DB 9; Length 388;
Best Local Similarity 80.8%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGACGATAGG 26
    ||||| ||||| ||||| |||||
Db 229 TCGAGCAACTTGTTCGGGAACGAAAGG 254

RESULT 40
US-10-027-632-107348
; Sequence 107348, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107348
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-107348

Query Match      60.0%; Score 18; DB 13; Length 597;
Best Local Similarity 80.8%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTTCAGGACGATAGGCT 28
    ||||| ||||| ||||| |||||
Db 247 GAGGGTCTTGTTCAGCAGCGGAGCCT 272
```

Search completed: January 7, 2005, 14:33:04  
Job time : 1291.26 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2005, 10:58:39 ; Search time 1791.05 Seconds  
(without alignments)  
610.363 Million cell updates/sec

Title: US-09-786-502A-8  
Perfect score: 30  
Sequence: 1 tcgaggatcttcaggacgtaggtgc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gss1: \*  
9: gb\_gss2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	70.0	608	5	BQ813384
2	21	70.0	650	6	CD488918
3	21	70.0	1178	5	BQ715651
4	20.8	69.3	310	9	CE517324
5	20.6	68.7	664	6	CA655944
6	20.4	68.0	736	9	CL155659
7	20.4	68.0	777	5	BP700670
8	20	66.7	320	6	CAL172592
9	20	66.7	473	8	BH051040
10	19.6	65.3	505	6	CA610501
11	19.6	65.3	548	8	BZ437621
12	19.6	65.3	636	2	BF046684
13	19.6	65.3	693	7	CF882870
14	19.6	65.3	711	8	AZ216274
15	19.6	65.3	749	6	CB909216
16	19.6	65.3	753	7	CF873769
17	19.6	65.3	772	8	BH504182
18	19.6	65.3	812	8	BH687374
19	19.6	65.3	823	6	CB903389
20	19.4	64.7	318	9	CL701378
21	19.4	64.7	328	1	AA486214
22	19.4	64.7	390	9	CE837920
23	19.4	64.7	395	6	CB773582
24	19.4	64.7	424	1	AJ706082

ORIGIN

25 19.4 64.7 466 9 CL702561  
c 26 19.4 64.7 555 8 AQ652867  
27 19.4 64.7 594 9 CL700756  
c 28 19.4 64.7 613 5 BQ401535  
c 29 19.4 64.7 616 1 AJ274168  
c 30 19.4 64.7 620 8 BZ519836  
c 31 19.4 64.7 621 7 CK086094  
32 19.4 64.7 627 9 CL703342  
c 33 19.4 64.7 633 9 AG125336  
c 34 19.4 64.7 634 9 CC688069  
c 35 19.4 64.7 636 8 BZ329488  
c 36 19.4 64.7 668 9 CE805733  
c 37 19.4 64.7 679 9 CC732980  
38 19.4 64.7 699 9 CL697382  
c 39 19.4 64.7 704 9 CL700732  
c 40 19.4 64.7 710 6 CB653308  
41 19.4 64.7 734 8 BZ696517  
42 19.4 64.7 741 4 BG281145  
43 19.4 64.7 952 5 BQ705973  
44 19.4 64.7 1201 8 BZ695091  
45 19.4 64.7 1560 3 BC022396

## ALIGNMENTS

RESULT 1  
BQ813384  
LOCUS BQ813384 608 bp mRNA linear EST 01-AUG-2002  
DEFINITION 1030036A05.v1 C. reinhardtii CC-1690, Deflagellation (normalized),  
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION BQ813384  
VERSION BQ813384.1 GI:22059341  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.  
REFERENCE 1 (bases 1 to 608)  
AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C.,  
Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.  
TITLES Analyses of the Chlamydomonas reinhardtii Genome: A Model  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1030  
JOURNAL Unpublished (2002)  
COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

## FEATURES

source  
Location/Qualifiers  
1..608  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Deflagellation  
(normalized), Lambda Zap II"  
/note="vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Deflagellation library, constructed by John Davies  
and Jeffrey McDermott, combines cDNAs from CC-1690 cells  
which had been re-synthesizing flagella for 15, 30 and 60  
min after being deflagellated by pH shock. PolyA mRNA was  
purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda Zap II  
(Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
Zap clones by superinfection with ExAssist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al., (1996) Genome Research 6: 791-806."

```
Query Match      70.0%; Score 21; DB 5; Length 608;
Best Local Similarity 82.8%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTCAGGAGCGATAGGCTGC 30
   ||||| ||||| ||||| ||||| ||||| |||||
DB 7 CGAGGATCTTGTCAGGAGCGCGCGGCGAGC 35

RESULT 2
CD488918/c
LOCUS      T16_B08 Teliospore Ustilago maydis cDNA 5', mRNA linear EST 29-AUG-2003
DEFINITION
ACCESSION  CD488918
VERSION     CD488918.1 GI:34331425
KEYWORDS    EST.
SOURCE      Ustilago maydis
ORGANISM    Ustilago maydis
            Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
            Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
REFERENCE   1 (bases 1 to 650)
AUTHORS     Sacadura,N.T. and Saville,B.J.
TITLE       Gene expression and EST analyses of Ustilago maydis germinating
            teliospores
JOURNAL     Fungal Genet. Biol. 40 (1), 47-64 (2003)
MEDLINE     22829673
PUBMED      12948513
COMMENT     Contact: Barry J. Saville
            Saville Lab
            University of Toronto
            3359 Mississauga Road North, Mississauga, ON, L5L 1C6, Canada
            Tel: 905 569 4702
            Fax: 905 828 3792
            Email: bsaville@utm.utoronto.ca
            Seq primer: M13 reverse primer (5' AAACAGCTATGACCATGTTCA 3').

FEATURES             source
    Location/Qualifiers
        1..650
            /organism="Ustilago maydis"
            /mol_type="mRNA"
            /strain="PBI/PB2"
            /db_xref="taxon:5270"
            /call_type="teliospore"
            /dev_stage="germinating teliospore"
            /lab_host="E. coli"
            /clone_lib="Teliospore"
            /note="Vector: pDNR-LIB; Site 1: SfiIA; Site 2: SfiIB;
            mRNA was extracted from germinating teliospores. cDNA was
            amplified by PCR and unidirectionally cloned into pDNR-LIB
            plasmid, with the use of Clontech's Creator SMART cDNA
            Library Construction Kit."

ORIGIN
Query Match      70.0%; Score 21; DB 6; Length 650;
Best Local Similarity 82.8%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTCAGGAGCGATAGGCTGC 30
   ||||| ||||| ||||| ||||| ||||| |||||
DB 623 CGAGGATCTTGTCATGCGCGCGGCTGC 595

RESULT 3
BQ715651/c
LOCUS      BQ715651 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6308233
DEFINITION  5', mRNA sequence.
ACCESSION  BQ715651
VERSION     BQ715651.1 GI:21854550
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1178)
NIH-MGC http://mgi.mgi.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13727 row: a column: 02
High quality sequence stop: 263.

FEATURES             source
    Location/Qualifiers
        1..1178
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:6308233"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 129"
            /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;
            Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.
            Primer: Oligo dt. Average insert size 2.2 Kb. Constructed
            by ResGen, Invitrogen Corp. Note: this is a NIH_MGC
            Library."

ORIGIN
Query Match      70.0%; Score 21; DB 5; Length 1178;
Best Local Similarity 82.8%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TCAGGATCTTGTCAGGAGCGATAGGCTG 29
   ||||| ||||| ||||| ||||| ||||| |||||
DB 944 TCAGGAGCTGTCAGGAGCGATTCGCTG 916

RESULT 4
CE517324
LOCUS      tigr-gss-dog-17000327420281 Dog Library Canis familiaris genomic,
DEFINITION  genomic survey sequence.
ACCESSION  CE517324
VERSION     CE517324.1 GI:36834105
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 310)
AUTHORS     Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE       The dog genome: survey sequencing and comparative analysis
JOURNAL     Science 301 (5641), 1898-1903 (2003)
MEDLINE     22875432
PUBMED      14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.

FEATURES             source
    Location/Qualifiers
        1..310
            /organism="Canis familiaris"
            /mol_type="genomic DNA"
            /strain="Standard Poodle"
```

```

/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match          69.3%; Score 20.8; DB 9; Length 310;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 GATCTTGTTCAGGAGCGATAGGCTG 29
||||| ||||| ||||| ||||| |||||
Db 1 GATCTTGGCAGGAGCGATAGGATG 24

RESULT 5
CA655944/c
LOCUS
DEFINITION
wlm0.pk0021.c4 wlm0 Triticum aestivum cDNA clone wlm0.pk0021.c4 5',
end, mRNA sequence.
CA655944
CA655944.1 GI:25234469
EST.
Triticum aestivum (bread wheat)
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
1 (bases 1 to 664)
Tingey, S. V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
Miao, G., Caraher, N. and Hanafey, M. K.
Dupont Wheat cDNA Sequence
Unpublished (2002)
Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
Location/Qualifiers
1..664
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Stephens"
/db_xref="taxon:4565"
/clone="wlm0.pk0021.c4"
/tissue_type="leaf"
/clone_lib="wlm0"
/notes="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) seedlings 0 hr after
inoculation with Erysiphe graminis f. sp tritici"

ORIGIN
Query Match          68.7%; Score 20.6; DB 6; Length 664;
Best Local Similarity 85.2%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CGAGGATCTGTTCAGGAGCGATAGGCT 28
||||| ||||| ||||| ||||| |||||
Db 550 CGAGGAGCCTGTTCGGAGCGATGGCT 524

RESULT 6
CL155659/c
LOCUS
DEFINITION
104.342.10782305.114 31473.017 Sorghum methylation-filtered library
(libid: 104) Sorghum bicolor genomic clone 10782305, genomic survey
sequence.
CL155659
CL155659.1 GI:40655675
GSS.

```

```

SOURCE
ORGANISM
Sorghum bicolor (sorghum)
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 736)
Budiman, M. A., Flick, E., Jones, J., Nunberg, A., Citek, R. W.,
Robbins, D., Rohlfing, T., Bradford, K., Fries, J., McHenry, J.,
Trani, L., Isak, A., Zimmerman, C., Lakey, N. and Bedell, J. A.
GeneThresher methylation filtered genomic sequences from Sorghum
bicolor
Unpublished (2004)
Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: 342 row: a column: 17
Seq primer: M13/pUC Forward
Class: shotgun
High quality sequence stop: 736.
Location/Qualifiers
1..736
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/cultivar="ATx623"
/db_xref="taxon:4558"
/clone="10782305"
/clone_lib="Sorghum methylation-filtered library (LibID:
104)"
/notes="Organ: leaf; Vector: pBSK(-); Site 1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBSK(-) vector
and electroporated into E. coli cells. This is a
methylation-filtered library."

ORIGIN
Query Match          68.0%; Score 20.4; DB 9; Length 736;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTCTCAGGAGCGATAGGCTGC 30
||||| ||||| ||||| ||||| |||||
Db 646 TCGAGGATCTCTCGGGGAAGATCGGCTGC 617

RESULT 7
BP700670/c
LOCUS
DEFINITION
BP700670 Osada Taira anterior neuroectoderm (ANE) PCS105 cDNA
library Xenopus laevis cDNA clone XL490h05ex 5', mRNA sequence.
BP700670
BP700670.1 GI:46049030
EST.
Xenopus laevis (African clawed frog)
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 777)
Osada, S., Kitayama, A., Ueno, N. and Taira, M.
Expression analysis of genes which are expressed in the anterior
neuroectoderm of Xenopus embryos
Unpublished (2004)
Contact: Masanori Taira
Department of Biological Sciences
Graduate School of Science, University of Tokyo; CREST, Japan
Science and Technology Corporation, Japan
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-03-5841-4434
Fax: 81-03-5841-4434

```





```

end, mRNA sequence.
CA610501.1 GI:25165663
EST.
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 505)
Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
DuPont Wheat cDNA Sequence
Unpublished (2002)
Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES
    source
        1..505
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /db_xref="taxon:4565"
            /clone="wrl.pk0117.c5"
            /tissue type="root"
            /clone lib="wrl"
            /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
            XhoI; Wheat (Triticum aestivum L.) root; 7 day old
            seedling, light grown"
ORIGIN
    Query Match          65.3%; Score 19.6; DB 6; Length 505;
    Best Local Similarity 84.8%; Pred.No. 6.6e+02;
    Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GAGGATCTTGTTCAGGAGCGATAGGCT 28
|||||
Db 157 GAGGATCTTGTTCAGGAGCGATAGGCT 182

RESULT 11
BZ437621/c
LOCUS
DEFINITION
    BZ437621 BO 1.6_2 KB tot Brassica oleracea genomic clone BONPJ41,
    genomic survey sequence.
ACCESSION
    BZ437621
VERSION
    BZ437621.1 GI:26691270
SOURCE
    Brassica oleracea
    ORGANISM
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
        rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
        1 (bases 1 to 548)
        Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
        Whole genome shotgun sequencing of Brassica oleracea
        Unpublished (2001)
        Other_GSSs: BONPJ41TR
        Contact: Chris Town
        TIGR
        9712 Medical Center Drive, Rockville, MD 20850, USA.
        Tel: 301-838-3523
        Fax: 301-838-0208
        Email: cdtown@tigr.org
        DNA is from a doubled haploid provided by Tom Osborn.
        Seq primer: TF
        Class: sheared ends.
FEATURES
    source
        1..548
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BONPJ41"
            /clone lib="BO 1.6_2 KB tot"
            /note="Vector: pHO51; Site 1: BstXI; 1.6-2 kb sheared
            total DNA inserted into pHO51 using BstXI linkers"
ORIGIN
    Query Match          65.3%; Score 19.6; DB 8; Length 548;
    Best Local Similarity 84.6%; Pred.No. 6.7e+02;
    Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AGGATCTTGTTCAGGAGCGATAGGCTG 29
|||||
Db 89 AGGATCTTGTTCAGGAGCGATAGGCTG 64

RESULT 12
BF046684
LOCUS
DEFINITION
    BF046684 636 bp mRNA linear EST 10-OCT-2000
    clone BP250020A10B3 Soares normalized bovine placenta Bos taurus cDNA
    sequence.
ACCESSION
    BF046684
VERSION
    BF046684.1 GI:10763728
KEYWORDS
    Bos taurus (cow)
SOURCE
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
        Bovinae; Bos.
        1 (bases 1 to 636)
        Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and
        Larson,J.H.
        Bovine ESTs
        Unpublished (2000)
        Contact: Lewin, H. A.
        W. M. Keck Center for Comparative and Functional Genomics
        University of Illinois at Urbana-Champaign
        340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
        61801, USA
        Tel: 217 333 5998
        Fax: 217 244 5617
        Email: h-lewin@uiuc.edu
        Funding for cattle EST sequencing was provided by the USDA National
        Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
        to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
        from Washington University Genome Center. Vector Trimmi g:
        Cross_match from Washington University Genome Center PHRAP suite.
        This sequence is vector free and at least 200 bp in length.
        PCR Primers
        FORWARD: TAATACGACTCACTATAGG
        BACKWARD: ATTACCCCTCACTAAG
        Insert Length: 636 Std Error: 0.00
        Plate: BP250020A10 row: B column: 3
        Seq primer: AGCGATAACAATTTCACACAGGA
        High quality sequence stop: 636.
        Location/Qualifiers
            1..636
                /organism="Bos taurus"
                /mol_type="mRNA"
                /db_xref="taxon:9913"
                /clone="BP250020A10B3"
                /sex="female"
                /lab host="DH10B"
                /clone lib="Soares normalized bovine placenta"
                /note="Organ: placenta; Vector: pT73Pac; Site 1: EcoRI;
                Site 2: NotI; The cDNA library was contributed by the
                Soares laboratory and it was constructed and normalized
                as described by Bonaldo, M.F., Lennon, G. and Soares,
                M.B. (1996), Genome Research 6(9): 791-806."
ORIGIN

```

```

Query Match      65.3%; Score 19.6; DB 2; Length 636;
Best Local Similarity 84.6%; Pred. No. 6.9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GGATCTTGTCTGAGGCGCATAGGCTGC 30
    ||||| ||||| ||||| ||||| |||||
Db 141 GCATCTTGGCAGGAGTGATAGCTGC 166

RESULT 13
CF882870/c
LOCUS
DEFINITION      CF882870.693 bp mRNA linear EST 31-OCT-2003
Hypocrea jecorina CDNA clone trico87xj04, Version 6 October 2003
ACCESSION      CF882870
VERSION
KEYWORDS
SOURCE
ORGANISM        Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
AUTHORS        Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
Ward, M. and Dean, R.A.
TITLE          Characterization of the protein processing and secretion pathways
in a comprehensive set of expressed sequence tags from Trichoderma
reesei
JOURNAL        FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
COMMENT        Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu
Seq primer: L1-P1 primer.
Location/Qualifiers
source
1..693
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/db_xref="taxon:51453"
/clone="trico87xj04"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/notes="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match      65.3%; Score 19.6; DB 7; Length 693;
Best Local Similarity 84.6%; Pred. No. 6.9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TCGAGGATCTTGTCTGAGGCGCATAGG 26
    ||||| ||||| ||||| ||||| |||||
Db 165 TCGAGGATCTTGTCTGAGGCGCATAGG 140

RESULT 14
AZ216274/c
LOCUS
DEFINITION      AZ216274.711 bp DNA linear GSS 09-JUN-2000
Sheared DNA-57B2.TR Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-57B2, genomic survey sequence.
ACCESSION      AZ216274
VERSION
KEYWORDS
SOURCE
ORGANISM        Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.
1 (bases 1 to 711)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leach, V., de Jong, P., Ullu, E., Melville, S.,
Donelson, J., Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other GSSs: Sheared DNA-57B2.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through Research Genetics, Alabama, USA. Sheared DNA
end sequences search page: http://www.tigr.org/tcdb/mbd/tbdb/.
Seq primer: M13-Reverse
Class: Shotgun
Location/Qualifiers
1..711
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-57B2"
/clone_lib="Sheared DNA"
/notes="Vector: pUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
Press, 1999)."

ORIGIN
Query Match      65.3%; Score 19.6; DB 8; Length 711;
Best Local Similarity 84.6%; Pred. No. 6.9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GAGGATCTTGTCTGAGGCGCATAGGCT 28
    ||||| ||||| ||||| ||||| |||||
Db 226 GAGGATCTTGTCTGAGGCGCATAGCT 201

RESULT 15
CB909216/c
LOCUS
DEFINITION      CB909216.749 bp mRNA linear EST 02-JUL-2003
trico87xj04 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone trico87xj04, mRNA sequence.
ACCESSION      CB909216
VERSION
KEYWORDS
SOURCE
ORGANISM        Hypocrea jecorina (anamorph: Trichoderma reesei)
Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 749)
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
22803314

```

12788920  
PUBMED  
COMMENT  
Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817  
Email: pforeman@genencor.com  
Seq primer: LR-F1 primer.  
Location/Qualifiers  
1. .749  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric087x104"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 3 april"  
/note="Vector: pREP3V; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."  
ORIGIN  
Query Match 65.3%; Score 19.6; DB 6; Length 749;  
Best Local Similarity 84.6%; Pred. No. 7e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0;  
QY 1 TCGAGGATCTTGTCTCAGGAGCGATAGG 26  
|||||  
Db 221 TCGAGGATCTTGTCTCAGGAGCGATAGG 196  
|||||  
RESULT 16  
CF873769/c  
LOCUS  
DEFINITION  
tric034xf13.bi T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric034xf13, mRNA sequence.  
CF873769  
CF873769.1 GI:38128451  
EST.  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocrocyetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 753)  
Diener.S.E., Dunn-Coleman.N., Foreman.P., Houfek.T.D.,  
Teunissen.P.J.M., van Solingen.P., Dankmeyer.L., Mitchell.T.K.,  
Ward.M. and Dean.R.A.  
Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei  
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LR-F1 primer.  
Location/Qualifiers  
1. .753  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric034xf13"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3V; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."  
ORIGIN  
Query Match 65.3%; Score 19.6; DB 6; Length 749;  
Best Local Similarity 84.6%; Pred. No. 7e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0;  
QY 1 TCGAGGATCTTGTCTCAGGAGCGATAGG 26  
|||||  
Db 221 TCGAGGATCTTGTCTCAGGAGCGATAGG 196  
|||||  
RESULT 16  
CF873769/c  
LOCUS  
DEFINITION  
tric034xf13.bi T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric034xf13, mRNA sequence.  
CF873769  
CF873769.1 GI:38128451  
EST.  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocrocyetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 753)  
Diener.S.E., Dunn-Coleman.N., Foreman.P., Houfek.T.D.,  
Teunissen.P.J.M., van Solingen.P., Dankmeyer.L., Mitchell.T.K.,  
Ward.M. and Dean.R.A.  
Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei  
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LR-F1 primer.  
Location/Qualifiers  
1. .753  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric034xf13"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3V; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."  
ORIGIN  
Query Match 65.3%; Score 19.6; DB 6; Length 749;  
Best Local Similarity 84.6%; Pred. No. 7e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0;  
QY 1 TCGAGGATCTTGTCTCAGGAGCGATAGG 26  
|||||  
Db 221 TCGAGGATCTTGTCTCAGGAGCGATAGG 196  
|||||  
RESULT 16  
CF873769/c  
LOCUS  
DEFINITION  
tric034xf13.bi T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric034xf13, mRNA sequence.  
CF873769  
CF873769.1 GI:38128451  
EST.  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocrocyetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 753)  
Diener.S.E., Dunn-Coleman.N., Foreman.P., Houfek.T.D.,  
Teunissen.P.J.M., van Solingen.P., Dankmeyer.L., Mitchell.T.K.,  
Ward.M. and Dean.R.A.  
Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei  
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LR-F1 primer.  
Location/Qualifiers  
1. .753  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric034xf13"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3V; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."  
ORIGIN  
Query Match 65.3%; Score 19.6; DB 6; Length 749;  
Best Local Similarity 84.6%; Pred. No. 7e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0;  
QY 1 TCGAGGATCTTGTCTCAGGAGCGATAGG 26  
|||||  
Db 221 TCGAGGATCTTGTCTCAGGAGCGATAGG 196  
|||||  
RESULT 16  
CF873769/c  
LOCUS  
DEFINITION  
tric034xf13.bi T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric034xf13, mRNA sequence.  
CF873769  
CF873769.1 GI:38128451  
EST.  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocrocyetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 753)  
Diener.S.E., Dunn-Coleman.N., Foreman.P., Houfek.T.D.,  
Teunissen.P.J.M., van Solingen.P., Dankmeyer.L., Mitchell.T.K.,  
Ward.M. and Dean.R.A.  
Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei  
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LR-F1 primer.  
Location/Qualifiers  
1. .753  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric034xf13"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3V; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."  
ORIGIN  
Query Match 65.3%; Score 19.6; DB 6; Length 749;  
Best Local Similarity 84.6%; Pred. No. 7e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0;  
QY 1 TCGAGGATCTTGTCTCAGGAGCGATAGG 26  
|||||  
Db 221 TCGAGGATCTTGTCTCAGGAGCGATAGG 196  
|||||  
RESULT 16  
CF873769/c  
LOCUS  
DEFINITION  
tric034xf13.bi T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric034xf13, mRNA sequence.  
CF873769  
CF873769.1 GI:38128451  
EST.  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocrocyetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 753)  
Diener.S.E., Dunn-Coleman.N., Foreman.P., Houfek.T.D.,  
Teunissen.P.J.M., van Solingen.P., Dankmeyer.L., Mitchell.T.K.,  
Ward.M. and Dean.R.A.  
Characterization of the protein processing and secretion pathways  
in a comprehensive set of expressed sequence tags from Trichoderma  
reesei  
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LR-F1 primer.  
Location/Qualifiers  
1. .753  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric034xf13"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3V; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."  
ORIGIN  
Query Match 65.3%; Score 19.6; DB 6; Length 749;  
Best Local Similarity 84.6%; Pred. No. 7e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0;  
QY 1 TCGAGGATCTTGTCTCAGGAGCGATAGG 26  
|||||  
Db 221 TCGAGGATCTTGT

TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other GSSs: BOHUX53TF  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

## FEATURES

source

Location/Qualifiers  
1. .812  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHUX53"  
/notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

## ORIGIN

Query Match 65.3%; Score 19.6; DB 8; Length 812;  
Best Local Similarity 84.6%; Pred. No. 7e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AGGATCTTGTTCAGGCGATAGCGTG 29

Db 394 AGGATCTTGTTCATGAAGAAGGCTG 369

## RESULT 19

CB903389/c

LOCUS trico34xf13 T.reesei mycelial culture, Version 3 april Hypocrea  
DEFINITION jecorina cDNA clone trico34xf13, mRNA sequence.

ACCESSION CB903389

VERSION CB903389.1 GI:30118047

KEYWORDS EST.

SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM

REFERENCE 1 (bases 1 to 823)  
Foreman P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,  
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,  
Olivares, H.A., Teunissen, P.J., Yao, J., and Ward, M.

Transcriptional regulation of biomass-degrading enzymes in the  
filamentous fungus Trichoderma reesei

J. Biol. Chem. 278 (34), 31988-31997 (2003)

JOURNAL

MEDLINE

PUBMED

COMMENT Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817  
Email: Pforeman@genencor.com

Seq primer: LT-Fi primer.

Location/Qualifiers

1. .823

/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/dev\_stage="mycelia"  
/clone lib="T. reesei  
/note="Vector: pREP3y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon

## FEATURES

source

Location/Qualifiers  
1. .823  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/dev\_stage="mycelia"  
/clone lib="T. reesei  
/note="Vector: pREP3y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon

## ORIGIN

Query Match 65.3%; Score 19.6; DB 6; Length 823;  
Best Local Similarity 84.6%; Pred. No. 7.1e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGCGATAGG 26

Db 120 TCGAGGATCTTGTTCAGGCGATAGG 95

## RESULT 20

CL701378

LOCUS

DEFINITION SP\_Ba0072B17.r SP\_Ba Sorghum propinquum genomic clone  
SP\_Ba0072B17 3', genomic survey sequence.

ACCESSION CL701378

VERSION CL701378.1 GI:50267653

KEYWORDS GSS.

SOURCE

ORGANISM

Sorghum propinquum  
Sorghum propinquum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 318)

Kim, H., Yu, Y., Missotski, M., Yost, D., Stum, D., Rao, K., Kudrna, D.,  
Muller, C., Soderlund, C., Bowers, J.E., Paterson, A.H. and Wing, R.

Physical mapping of the sorghum genome

Unpublished (2004)

Contact: Rod A. Wing

Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: http://genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0072 row: B column: 17

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

Location/Qualifiers

1. .318

/organism="Sorghum propinquum"

/mol\_type="genomic DNA"

/db\_xref="taxon:132711"

/clone="SP\_Ba0072B17"

/clone lib="SP\_Ba"

/note="Vector: pBelobAC11; Site 1: HindIII; Site 2:  
HindIII; Paterson lab BAC library (HindIII)"

## ORIGIN

Query Match 64.7%; Score 19.4; DB 9; Length 318;  
Best Local Similarity 79.3%; Pred. No. 7.6e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTTCAGGCGATAGGCTGC 30

Db 108 CGACTATCTTTTCAGGACCAATGGCTGC 136

## RESULT 21

AA486214

LOCUS

DEFINITION ab35d07.s1 StrataGene Hela cell s3 937216 Homo sapiens cDNA clone  
IMAGE:842797 3', similar to gb:S71381 PROTEASOME BETA CHAIN  
(HUMAN), mRNA sequence.

ACCESSION AA486214

VERSION AA486214.1 GI:2216430

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 328)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, P., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished (1997)

COMMENT Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Insert Length: 1065 Std Error: 0.00

Seq primer: -41m13 fwd. ET from Amersham

High quality sequence stop: 1.

FEATURES

source

1..328

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:842797"

/sex="female"

/dev\_stage="Hela S3 cell line"

/lab\_host="SOLR (kanamycin resistant)"

/clone\_lib="Stratagene HeLa cell s3 937216"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTCAGTGTGTGTGTGTGTGTGTGT 3'"

ORIGIN

Query Match 64.7%; Score 19.4; DB 1; Length 328;

Best Local Similarity 79.3%; Pred. No. 7.7e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTG 29

DB 299 TCGCAGACTTTCGGGAGCGAGAGGCTG 327

RESULT 22

CEB37920/c

LOCUS CEB37920 390 bp DNA linear GSS 30-SEP-2003

DEFINITION tigr-gss-dog-17000332602409 Dog Library Canis familiaris genomic, genomic survey sequence.

ACCESSION CEB37920

VERSION CEB37920

KEYWORDS GSS.

SOURCE Canis familiaris (dog)

ORGANISM Canis familiaris

REFERENCE 1 (bases 1 to 390)

AUTHORS Kirschner, E.P., Bafna V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Frazer, C.M. and Venter, J.C.

TITLE The dog genome: survey sequencing and comparative analysis

JOURNAL Science 301 (5641), 1898-1903 (2003)

MEDLINE 22875432

PUBMED 14512627

COMMENT Contact: Kirschner EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 328)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, P., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished (1997)

COMMENT Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Insert Length: 1065 Std Error: 0.00

Seq primer: -41m13 fwd. ET from Amersham

High quality sequence stop: 1.

FEATURES

source

1..328

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:842797"

/sex="female"

/dev\_stage="Hela S3 cell line"

/lab\_host="SOLR (kanamycin resistant)"

/clone\_lib="Stratagene HeLa cell s3 937216"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTCAGTGTGTGTGTGTGTGTGTGT 3'"

ORIGIN

Query Match 64.7%; Score 19.4; DB 1; Length 328;

Best Local Similarity 79.3%; Pred. No. 7.7e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTG 29

DB 299 TCGCAGACTTTCGGGAGCGAGAGGCTG 327

RESULT 22

CEB37920/c

LOCUS CEB37920 390 bp DNA linear GSS 30-SEP-2003

DEFINITION tigr-gss-dog-17000332602409 Dog Library Canis familiaris genomic, genomic survey sequence.

ACCESSION CEB37920

VERSION CEB37920

KEYWORDS GSS.

SOURCE Canis familiaris (dog)

ORGANISM Canis familiaris

REFERENCE 1 (bases 1 to 390)

AUTHORS Kirschner, E.P., Bafna V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Frazer, C.M. and Venter, J.C.

TITLE The dog genome: survey sequencing and comparative analysis

JOURNAL Science 301 (5641), 1898-1903 (2003)

MEDLINE 22875432

PUBMED 14512627

COMMENT Contact: Kirschner EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 328)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, P., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished (1997)

COMMENT Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Insert Length: 1065 Std Error: 0.00

Seq primer: -41m13 fwd. ET from Amersham

High quality sequence stop: 1.

FEATURES

source

1..328

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:842797"

/sex="female"

/dev\_stage="Hela S3 cell line"

/lab\_host="SOLR (kanamycin resistant)"

/clone\_lib="Stratagene HeLa cell s3 937216"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTCAGTGTGTGTGTGTGTGTGTGT 3'"

ORIGIN

Query Match 64.7%; Score 19.4; DB 1; Length 328;

Best Local Similarity 79.3%; Pred. No. 7.7e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTG 29

DB 299 TCGCAGACTTTCGGGAGCGAGAGGCTG 327

RESULT 22

CEB37920/c

LOCUS CEB37920 390 bp DNA linear GSS 30-SEP-2003

DEFINITION tigr-gss-dog-17000332602409 Dog Library Canis familiaris genomic, genomic survey sequence.

ACCESSION CEB37920

VERSION CEB37920

KEYWORDS GSS.

SOURCE Canis familiaris (dog)

ORGANISM Canis familiaris

REFERENCE 1 (bases 1 to 390)

AUTHORS Kirschner, E.P., Bafna V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Frazer, C.M. and Venter, J.C.

TITLE The dog genome: survey sequencing and comparative analysis

JOURNAL Science 301 (5641), 1898-1903 (2003)

MEDLINE 22875432

PUBMED 14512627

COMMENT Contact: Kirschner EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 328)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, P., Theising, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished (1997)

COMMENT Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Insert Length: 1065 Std Error: 0.00

Seq primer: -41m13 fwd. ET from Amersham

High quality sequence stop: 1.

FEATURES

source

1..328

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:842797"

/sex="female"

/dev\_stage="Hela S3 cell line"</

```

KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 424)
Depitta, C., Tomblan, L., Kronnie, G., Romualdi, C., Vitulo, N.,
Basso, G., and Lanfranchi, G.
TITLE A leukemia-enriched cDNA microarray platform identified new
transcripts with relevance to the biology of leukemias
JOURNAL Unpublished (2004)
COMMENT Contact: Depitta C
Biology and CRIBI
University of Padova
Via U. Bassi, 58/B, 35131, ITALY.

FEATURES
source
1..424
Location/Qualifiers
/mol_type="Homo sapiens"
/db_xref="taxon:9606"
/clone="BLPD01229"
/tissue_type="bone marrow"
/clone_lib="BLPD01"
/notes="Caucasian"

ORIGIN
Query Match 64.7%; Score 19.4; DB 1; Length 424;
Best Local Similarity 79.3%; Pred. No. 7.9e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTG 29
|||||
Db 114 TCGAGGTTCTTCTGAGGACGATCGCTG 142

RESULT 25
CL702561
LOCUS CL702561 466 bp DNA linear GSS 13-JUL-2004
DEFINITION SP_Ba0085N23.f SP_Ba Sorghum propinquum genomic clone
SP_Ba0085N23 5', genomic survey sequence.
ACCESSION CL702561
VERSION CL702561
KEYWORDS GSS.
SOURCE Sorghum propinquum
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
AUTHORS 1 (bases 1 to 466)
Kim, H., Yu, Y., Wissotski, M., Yost, D., Stum, D., Rao, K., Kudrna, D.,
Muller, C., Soderlund, C., Bowers, J.E., Paterson, A.H. and Wing, R.
TITLE Physical mapping of the sorghum genome
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0085 row: N column: 23
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.

FEATURES
source
1..466
Location/Qualifiers
/mol_type="Sorghum propinquum"
/mol_type="genomic DNA"
/db_xref="taxon:132711"
/clone="SP_Ba0085N23"
/clone_lib="SP_Ba"

/note="Vector: pBeoBAC11; Site 1: HindIII; Site 2: HindIII; Paterson lab BAC library (HindIII)"
HindIII; Paterson lab BAC library (HindIII)"

Query Match 64.7%; Score 19.4; DB 9; Length 466;
Best Local Similarity 79.3%; Pred. No. 8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 CGAGGATCTTGTTCAGGAGCGATAGGCTG 30
|||||
Db 166 CGACTATCTTTCAGGATCAATTGGCTG 194

RESULT 26
AQ652667/c
LOCUS AQ652667/c 555 bp DNA linear GSS 22-JUN-1999
DEFINITION Sheared DNA-17N21-TR Sheared DNA Trypanosoma brucei genomic clone
AQ652667
VERSION AQ652667
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
AUTHORS 1 (bases 1 to 555)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S.,
Donelson, J., Fraser, C. and Adams, M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other GSSs: Sheared DNA-17N21.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tldb/mdb/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.

FEATURES
source
1..555
Location/Qualifiers
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-17N21"
/clone_lib="Sheared DNA"

/note="Vector: pUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
Press, 1999)."

Query Match 64.7%; Score 19.4; DB 8; Length 555;
Best Local Similarity 79.3%; Pred. No. 8.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTG 29
|||||

```



```

Qy 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTG 29
    ||||| ||| ||| ||||| |||||
Db 515 TCGAGGATATAGTCGGAAGCGGAGGCTG 487

RESULT 30
BZ519836/c
LOCUS BOMQD55TF BO_2_3 KB Brassica oleracea genomic clone BOMQD55, linear GSS 16-DEC-2002
DEFINITION genomic survey sequence.
ACCESSION BZ519836
VERSION BZ519836.1 GI:27052121
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 620)
AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOMQD55TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.
FEATURES
    source
        1..620
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="Tol000DH3"
        /db_xref="taxon:3712"
        /clone="BOMQD55"
        /notes="Vector: pHOS1; site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN
Query Match 64.7%; Score 19.4; DB 8; Length 620;
Best Local Similarity 79.3%; Pred. No. 8.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 CGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    ||||| ||||| ||||| |||||
Db 590 CCAGCTTCTTGTGAGGAACGATAGCCTGC 562

RESULT 31
LOCUS CK086094
DEFINITION RG11_F05 Cucurbit leaf Cucumis sativus cDNA, mRNA linear EST 01-DEC-2003
ACCESSION CK086094
VERSION CK086094.1 GI:38571154
KEYWORDS EST.
SOURCE Cucumis sativus (cucumber)
ORGANISM Cucumis sativus
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
REFERENCE 1 (bases 1 to 621)
AUTHORS Grumet, R. and McGrath, M.
TITLE Development of genomic tools for cucumber (Cucumis sativus L.)
JOURNAL Unpublished (2003)
COMMENT Contact: Rebecca Grumet
Rebecca Grumet
Michigan State University
Horticulture Department, Michigan State University, East Lansing,

MI 48824, USA
Tel: 517 353 0890
Fax: 517 355 5191 x431
Email: grumet@msu.edu
Plate: RG11 row: F column: 05.
Location/Qualifiers
    1..621
    /organism="Cucumis sativus"
    /mol_type="mRNA"
    /strain="Straight 8"
    /db_xref="taxon:3659"
    /sex="monoecious"
    /clone_lib="Cucumber leaf"
    /note="Vector: PAD-GAL4; Site_1: EcoRI; Site_2: XhoI"
ORIGIN
Query Match 64.7%; Score 19.4; DB 7; Length 621;
Best Local Similarity 79.3%; Pred. No. 8.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TCGAGGATCTTGTTCAGGAGCGATAGGCTG 29
    ||||| ||||| ||||| |||||
Db 178 TCGAAGATCTTGACACGATAGTAGGCTG 206

RESULT 32
LOCUS CL703342
DEFINITION SP_Ba0095F13 r SP_Ba Sorghum propinquum genomic clone
ACCESSION CL703342
VERSION CL703342.1 GI:50269617
KEYWORDS GSS.
SOURCE Sorghum propinquum
ORGANISM Sorghum propinquum
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 627)
AUTHORS Kim, H., Yu, Y., Wissotski, M., Yost, D., Stum, D., Rao, K., Kudrna, D.,
Muller, C., Soderlund, C., Bowers, J.E., Paterson, A.H. and Wing, R.
TITLE Physical mapping of the sorghum genome
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0095 row: F column: 13
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
Location/Qualifiers
    1..627
    /organism="Sorghum propinquum"
    /mol_type="genomic DNA"
    /db_xref="taxon:132711"
    /clone="SP_Ba0095F13"
    /clone_lib="SP_Ba"
    /note="Vector: pBelOAC11; Site 1: HindIII; Site 2: HindIII; Paterson lab BAC library (HindIII)"
ORIGIN
Query Match 64.7%; Score 19.4; DB 9; Length 627;
Best Local Similarity 79.3%; Pred. No. 8.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 CGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    ||||| ||||| ||||| |||||

```



```

Db      173 CGACTATCTTTTCAGGATCAATGGCTGC 201

RESULT 33
LOCUS   AG125336/c
DEFINITION Pan troglodytes DNA, clone: PTB-135124.R, genomic survey sequence.
ACCESSION AG125336
VERSION   AG125336.1 GI:16654501
KEYWORDS GSS.
SOURCE   Pan troglodytes (chimpanzee)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
AUTHORS  Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
          Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE    BAC end sequences of Library PTB
JOURNAL  Unpublished
AUTHORS  2 (bases 1 to 633)
          Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
          Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE    Direct Submission
JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail: chimpes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
          Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT  Clones are derived from the chimpanzee BAC library PTB This BAC end
          was generated during the R&D process and may have higher chance of
          clone tracking errors.
          PRIMERS
          Sequencing: M13Rev
LIBRARY
Vector   : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
source   Location/Qualifiers
    1..633
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /clone="PTB-135124.R"
        /sex="male"
        /cell_type="lymphoblast"
        /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match      64.7%; Score 19.4; DB 9; Length 633;
Best Local Similarity 79.3%; Pred. No. 8.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1 TCGAGGATCTTGTTCAGGAGCGATGGCTG 29
      |||||
Db      589 TTGAGGATCTTGTTCAGGAGCGATGGCTG 561

RESULT 34
LOCUS   CC688069/c
DEFINITION CGVEP58TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMEMa0537020,
          genomic survey sequence.
ACCESSION CC688069
VERSION   CC688069.1 GI:32092845
KEYWORDS GSS.
SOURCE   Zea mays
ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoidae; Andropogoneae; Zea.
REFERENCE
AUTHORS  Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
          Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
          Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
          Consortium for Maize Genomics
          Unpublished (2002)
          Contact: Cathy Whitelaw
          TIGR
          9712 Medical Center Drive, Rockville, MD 20850, USA
          Tel: 301-838-5843
          Fax: 301-838-0208
          Email: whitelaw@tigr.org
          Seq primer: TF
          Class: sheared ends.
          Location/Qualifiers
            1..634
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="taxon:4577"
                /clone="ZMMEMa0537020"
                /clone_lib="ZM_0.7_1.5_KB"
                /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
                methylation filtered genomic DNA library"
ORIGIN
Query Match      64.7%; Score 19.4; DB 9; Length 634;
Best Local Similarity 79.3%; Pred. No. 8.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

BZ329488      636 bp      DNA      linear      GSS 06-NOV-2002
hv86c01.b1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum
bicolor genomic clone hv86c01 5', genomic survey sequence.
BZ329488
VERSION   BZ329488.1 GI:24715306
KEYWORDS GSS.
SOURCE   Sorghum bicolor (sorghum)
ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE
AUTHORS  Rabinowicz, P.D., O'Shaughnessy, A.L., Balijs, V., Dedhia, N.,
          Katzenburger, F., King, L., Miller, B., Muller, S., Nascimeto, L.,
          Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.,
          Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
          Unpublished (2002)
          Contact: W. Richard McCombie
          Lita Annenberg Hazen Genome Sequencing Center
          Cold Spring Harbor Laboratory
          PO Box 100, Cold Spring Harbor, NY 11724, USA
          Tel: 516 367 8884
          Fax: 516 367 8874
          Email: mcombie@cshl.org
          Plate: hv86 row: c column: 01
          Seq primer: -21M13UnivFwd
          Class: shotgun
          High quality sequence stop: 636.
          Location/Qualifiers
            1..636
                /organism="Sorghum bicolor"
                /mol_type="genomic DNA"
                /db_xref="taxon:4558"
                /clone="hv86c01"
                /lab_host="JM107 or DH5a"
                /clone_lib="WGS-SbicolorF (JM107 adapted methyl filtered)"
                /note="Site 1: Xba I; Site 2: Xba I; The vector was
                digested with XbaI and one nucleotide was added by fill in
                in the recessive 3' end. The genomic DNA was nebulized,"
          TITLE
          JOURNAL
          COMMENT
FEATURES
source   Location/Qualifiers
    1..636
        /organism="Sorghum bicolor"
        /mol_type="genomic DNA"
        /db_xref="taxon:4558"
        /clone="hv86c01"
        /lab_host="JM107 or DH5a"
        /clone_lib="WGS-SbicolorF (JM107 adapted methyl filtered)"
        /note="Site 1: Xba I; Site 2: Xba I; The vector was
        digested with XbaI and one nucleotide was added by fill in
        in the recessive 3' end. The genomic DNA was nebulized,"
    
```



```
/organism="Sorghum propinquum"
/mol_type="genomic DNA"
/db_xref="taxon:132711"
/clone="SP_Ba0010H14"
/clone_lib="SP_Ba"
/notes="Vector: pBeloBAC11; Site 1: HindIII; Site 2:
HindIII; Paterson lab BAC library (HindIII)"

ORIGIN
Query Match      64.7%; Score 19.4; DB 9; Length 699;
Best Local Similarity 79.3%; Pred. No. 8.5e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    ||| ||||| ||||| ||||| ||||| |||||
Db 108 CGACTATCTTTTCAGGATCAATTGGCTGC 136

RESULT 39
CL700732
LOCUS CL700732 704 bp DNA linear GSS 13-JUL-2004
DEFINITION SP_Ba0064L21.f SP_Ba Sorghum propinquum genomic clone
SOURCE SP_Ba0064L21 5', genomic survey sequence.
ACCESSION CL700732
VERSION CL700732.1 GI:50267007
KEYWORDS GSS.
SOURCE Sorghum propinquum
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 704)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Kudrna,D.,
Muller,C., Soderlund,C., Bowers,J.E., Paterson,A.H. and Wing,R.
TITLE Physical mapping of the sorghum genome
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0064 row: L column: 21
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES
    source
    Location/Qualifiers
        1..704
        /organism="Sorghum propinquum"
        /mol_type="genomic DNA"
        /db_xref="taxon:132711"
        /clone="SP_Ba0064L21"
        /clone_lib="SP_Ba"
        /notes="Vector: pBeloBAC11; Site 1: HindIII; Site 2:
        HindIII; Paterson lab BAC library (HindIII)"

ORIGIN
Query Match      64.7%; Score 19.4; DB 9; Length 704;
Best Local Similarity 79.3%; Pred. No. 8.5e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    ||| ||||| ||||| ||||| ||||| |||||
Db 152 CGAATATCTTTTCAGGATCAATTGGCTGC 180

RESULT 40
CB653308/c
LOCUS CB653308 710 bp mRNA linear EST 09-APR-2003
DEFINITION OSJNEC04D10.r OSJNEC Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEC04D10 3', mRNA sequence.
CB653308
VERSION CB653308.1 GI:29657033
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.
REFERENCE 1 (bases 1 to 710)
AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: D column: 10
Seq primer: gga aac agc tat gac cat g.
FEATURES
    source
    Location/Qualifiers
        1..710
        /organism="Oryza sativa (japonica cultivar-group)"
        /mol_type="mRNA"
        /cultivar="Nipponbare"
        /db_xref="taxon:39947"
        /clone="OSJNEC04D10"
        /tissue_type="Leaf"
        /dev_stage="3 week"
        /lab_host="DH10B"
        /clone_lib="OSJNEC"
        /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
        XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN
Query Match      64.7%; Score 19.4; DB 6; Length 710;
Best Local Similarity 79.3%; Pred. No. 8.5e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CGAGGATCTTGTTCAGGAGCGATAGGCTGC 30
    ||| ||||| ||||| ||||| ||||| |||||
Db 660 CCAGGATCTCTCAGGAGCGAGAGCTGC 632

Search completed: January 7, 2005, 12:37:44
Job time : 1796.05 secs
```

